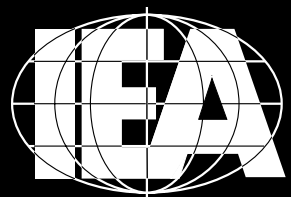


# **Technical Standards for IEA Studies: An Annotated Bibliography**

**Kelvin D. Gregory  
and Michael O. Martin**



International Association for  
the Evaluation of Educational  
Achievement

# Technical Standards for IEA Studies: An Annotated Bibliography

Kelvin D. Gregory and Michael O. Martin

September 2001



International Association for the  
Evaluation of Educational Achievement

© International Association for the Evaluation of Educational Achievement 2001  
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

ISBN 90 806643 1 6

Copies of *Technical Standards for IEA Studies: An Annotated Bibliography* can be obtained from:

IEA Secretariat

Herengracht 487

1017 BT Amsterdam

The Netherlands

Telephone +31 20 625 3625

Fax +31 20 420 7136

Email: [Department@IEA.nl](mailto:Department@IEA.nl)

Website: <http://www.iea.nl/>

*Text edited by:* Paula Wagemaker Editorial Services, Christchurch, New Zealand

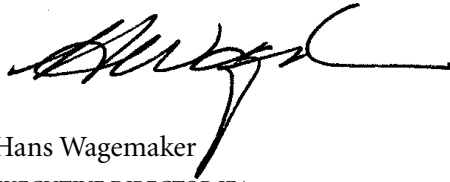
*Designed and desktop published by:* Becky Bliss Design and Production, Wellington, New Zealand

*Printed by:* Freestyle Artworks Ltd, Wellington, New Zealand

# Foreword

As part of its commitment to excellence and as a means of providing guidance to researchers involved in IEA studies, the Technical Executive Group of the International Association for the Evaluation of Educational Achievement produced *Technical Standards for IEA Studies*, a document that was published in 1999. *Technical Standards: An Annotated Bibliography* is intended to augment that publication and provide researchers with a guide to some of the more significant resource materials that are used when designing and completing our major studies. This document will be updated at regular intervals and will be available in its most current form from the IEA web site.

I would like to thank all those who contributed to this document by providing the details of the resources, especially the members of the IEA Technical Executive Group. I thank in particular Kelvin Gregory and Michael O. Martin for taking the major responsibility for ensuring the completion of this publication.

A handwritten signature in black ink, appearing to read 'Hans Wagemaker', with a long horizontal stroke extending to the right.

Hans Wagemaker  
EXECUTIVE DIRECTOR IEA



# Contents

<b>Foreword</b>	3
<b>Introduction</b>	7
<b>Structure of the Standards</b>	8
<b>Organization of the Bibliography</b>	10
<b>Sources for Individual Standards</b>	
<i>Designing, Managing and Implementing IEA Studies</i>	16
Initial planning of an IEA study	16
Choosing an international coordinating center	20
Formulating and refining study questions	22
Designing an IEA study	24
Developing a sampling plan	26
Choosing data collection methods	30
Developing a quality assurance program	31
Preparing an analysis plan	33
<i>Developing Data Collection Instruments</i>	36
Developing assessment frameworks and conceptual models	36
Developing specifications for tests and questionnaires	41
Test development	45
Questionnaire development	48
Translations and verifying translations	51
Field testing data collection instruments and procedures	53
<i>Data Collection and Processing</i>	55
Drawing a sample	55
Planning for data collection	57
Selecting and training data collection staff	58
Minimizing response burden and non-response	59
Implementing data collection quality control procedures	60
Documenting national data collection	62
Planning for data preparation and processing	63
Preparing data for processing and checking data	65
Documenting data-processing activities	67

<i>Analyzing Data and Reporting Results</i>	69
Developing sampling weights	69
Reporting sampling and non-sampling errors	70
Validating constructs and scales for analysis	72
Presenting findings	74
Reviewing the primary reports of study findings	78
Releasing data	79
Preparing technical reports and documentation	81



# Introduction

The International Association for the Evaluation of Educational Achievement (IEA) has as its mission to improve education through international comparative studies of student achievement that focus on educational policies and practices so as to enhance learning in systems of education all around the world. In furtherance of this mission, IEA has been conducting international studies for almost 50 years, and, through its network of participating countries, has accumulated a large reservoir of knowledge and experience in the field of international student assessment. In 1999, with support from the United States National Center for Education Statistics, IEA published its *Technical Standards for IEA Studies* (Martin, Rust & Adams, 1999), which sought to present in a brief volume a set of standards for good practice that would serve both as a yardstick against which to evaluate IEA studies and as a guide to good practice for future studies.

At the time it published the *Standards*, IEA committed itself to producing an annotated bibliography of reference works and other publications that would provide information and guidance to readers interested in learning more about the methodology that underpins IEA studies. The present volume is intended to fulfill that commitment. The bibliography begins with a brief summary of the standards and lists a set of published sources that are relevant to the standards in general. This is followed by a statement of each standard and a list of sources especially relevant to that standard. The number of references varies from standard to standard, because some of the areas addressed by the standards have received more attention in the literature than have others. Where possible, reference is made to technical documentation from existing IEA studies, as examples of the current state of practice. Much of this documentation is from the Third International Mathematics and Science Study (TIMSS), the largest of the IEA studies.

Although every effort was made in the time available to include a wide range of up-to-date sources in the bibliography, the selection naturally reflects to some degree the background and interests of the authors. However, IEA plans that the bibliography will be a 'living document', maintained on the IEA web site and updated and extended periodically.





# Structure of the Standards

The IEA standards are designed to address the major issues that have confronted IEA studies as they have been conducted over the years. In particular, the standards relating to IEA studies usually acknowledge the following: the measurement of student achievement in school subjects is a fundamental objective; data are collected by means of survey samples conducted in school settings; data are collected simultaneously in a large number of countries by national teams using internationally agreed instruments and following internationally agreed procedures; and the study is managed and coordinated by an international team. Reflecting this reality, the standards are grouped into four areas:

1. The design, management, operation and quality control of international studies.
2. The construction of instruments for measuring student achievement and questionnaires for collecting background information from students, teachers and schools.
3. Data collection by means of surveys in school settings.
4. Data processing, analysis and reporting.

Within each of the four areas there are six to nine individual standards, each with a brief statement of its purpose, and a statement that describes the standard itself. Each standard includes a set of guidelines to be followed to ensure that the standard will be met, and sometimes a checklist to help monitor the implementation of the standard. Standards have been developed in the following areas:

## **Standards for designing, managing and implementing IEA studies**

- Initial planning of an IEA study
- Choosing an international coordinating center
- Formulating and refining study questions
- Designing an IEA study
- Developing a sampling plan
- Choosing data collection methods
- Developing a quality assurance program
- Preparing an analysis plan.



## **Standards for developing data collection instruments**

- Developing assessment frameworks and conceptual models
- Developing specifications for tests and questionnaires
- Test development
- Questionnaire development
- Translations and verifying translations
- Field testing data collection instruments and procedures.

## **Standards for data collection and processing**

- Drawing a sample
- Planning for data collection
- Selecting and training data collection staff
- Minimizing response burden and non-response
- Implementing data collection quality control procedures
- Documenting national data collection
- Planning for data preparation and processing
- Preparing data for processing and checking data
- Documenting data-processing activities.

## **Standards for analyzing data and reporting results**

- Developing sampling weights
- Reporting sampling and non-sampling errors
- Validating constructs and scales for analysis
- Presenting findings
- Reviewing the primary reports of study findings
- Releasing data
- Preparing technical reports and documentation.



# Organization of the Bibliography

IEA studies are rooted in the traditions, philosophies and methodologies of educational research. As such, any bibliography that focuses solely on individual standards risks omitting much of the larger picture. Consequently, this volume begins with a selection of general sources that represents views and opinions from across the broad range of educational research. These sources address in a general way the concerns of all IEA studies, and since they include many excellent training, refresher and reference texts, they are good starting points for readers seeking a broad overview of the major issues. These general sources often are relevant to a number of individual standards. In some instances, where a general reference is especially pertinent, it is also referenced under the specific standard. For example, Keeves' volume *Educational Research, Methodology, and Measurement: An International Perspective* is referenced in this section and also under a number of specific standards.

Following the general sources, each standard is presented in turn, with a statement of its purpose, the standard itself, a summary of the guidelines provided in the *Standards* for meeting the standard, and a list of sources pertinent to the standard.

## General Sources

Black, T. R. (1999) *Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics*. London: Sage Publications.

This introductory textbook provides a comprehensive and integrated approach to using quantitative methods in the social sciences. Its primary focus is on the design and execution of research. Key topics such as planning, sampling, the design of measuring instruments, choice of statistical textbook and interpretation of results are examined within the context of the research process.

Bradley, W. J. and Schaefer, K. C. (1998) *The Uses and Misuses of Data and Models: The Mathematization of the Human Sciences*. Thousand Oaks, CA: Sage Publications.

This advanced, graduate-level textbook develops principles that can guide the use of data and models in the human sciences. The textbook offers information of general relevance to the standards. For example, the section 'Problems in Gathering Human Information' (pp. 9–14) highlights several sets of problems, namely those arising from assumptions inherent in a methodology originally developed for the natural sciences, those arising from the fact that data and models have the potential to provide great social benefits, and those associated with seeking an algorithmic explanation of phenomena. The last four pages of the textbook (pp. 190–93) provide an excellent summary of the book's main arguments, for example, 'Analysts must carefully represent the completeness and relevance of their work, along with the limitations of the data



and predictions' (p. 190). The textbook serves to reinforce central tenets of the IEA standards.

Cresswell, J. W. (1994) *Research Design: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.

This university textbook provides the basis for deciding whether to use a qualitative or quantitative approach for the design of a research study and how to write up the results of a study for a journal article or dissertation. The author offers a guide to the major design decisions, such as deciding upon a paradigm, stating the purpose of the study, identifying the research questions and hypotheses, using theory, and defining and stating the significance of the study.

Fink, A. (1995) *The Survey Handbook: The Survey Kit* (Vol. 1). Thousand Oaks, CA: Sage Publications.

This book provides an introduction to the skills and resources necessary for conducting a survey. Fink describes four different types of survey instruments, details the activities that constitute a 'typical' survey, and demonstrates how to organize surveys and estimate their costs. In addition, she discusses useful techniques such as what types of questions to ask in pilot tests, and how to execute a management plan and budget for a survey. This volume is the first in a series of easy-to-read, basic survey books. The series is useful for its checklists and broad coverage of survey issues, and is suitable for undergraduates.

Fraenkel, J. R. and Wallen, N. E. (2000) *How to Design and Evaluate Research in Education* (4th ed.). New York: McGraw-Hill.

This textbook provides the basic information needed to understand the research process, and seeks to enable people to design and conduct their own research. Suited to the graduate student, the book provides a comprehensive step-by-step guide to educational research. Chapters of this textbook are also referenced under specific IEA standards.

Gall, M. D., Borg, W. R. and Gall, J. P. (1996) *Educational Research: An Introduction* (6th ed.). New York: Longman.

Intended for university students, this textbook provides a comprehensive introduction to the major research methods and types of data analyses used in educational inquiry.

Gay, L. R. and Airasian, P. (1996) *Educational Research: Competencies for Analysis and Application* (6th ed.). Upper Saddle River, NJ: Merrill.

This textbook offers a comprehensive coverage of introductory concepts in educational research. Chapters include 'Selection and Definition of a Problem', 'Preparation and Evaluation of a Research Plan', 'Selection of a Sample', 'Selection of Measuring Instruments' and 'Preparation of a Research Report'.

Keeves, J. P. (Ed.) (1997) *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.). Cambridge, UK: Pergamon.

This book contains articles that relate directly to many of the IEA standards. The book represents perhaps the most complete and up-to-date single source of information pertaining to the standards and would serve as an invaluable main reference. Most of the articles are written by recognized leaders in the field.



Krathwohl, D. R. (1998) *Methods of Educational and Social Science Research: An Integrated Approach* (2nd ed.). New York: Longman.

This graduate-level textbook provides an integrated framework for research. Throughout the book, Krathwohl refers to scientific research although the textbook is specifically written for social and behavioral science research. The textbook presents a very thorough treatment of the variety of research methods commonly used in social and behavioral science research.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

Designed for people who are new to research and planning, this book presents advice on how to undertake a research project. The textbook is organized around eight operational steps that constitute the research process: formulating a research problem; conceptualizing a research design; constructing an instrument for data collection; selecting a sample; writing a research proposal; collecting data; processing data; and writing a research report.

Martin, M. O., Rust, K. and Adams, R. J. (Eds.), with contributions from Caldwell, N., Foy, P., Martin, M. O., Mullis, I. V. S. and Sibberns, S. (1999) *Technical Standards for IEA Studies*. Amsterdam: IEA.

This publication presents the technical standards developed by IEA, including a rationale for the approach adopted. A statement of each standard is presented, along with a set of guidelines to be followed so that the standard will be met. For some standards, a checklist to help monitor the implementation of the standard is also provided. The annotated bibliography is intended to complement the standards by providing readers with reference sources that provide further information about the methodological underpinnings of IEA studies.

Mason, E. J. and Bramble, W. J. (1997) *Research in Education and Behavioral Sciences: Concepts and Methods*. New York: McGraw-Hill.

Graduate students are introduced to the nature of educational and behavioral science research, the tools and methods they need to conduct it, how to interpret and report the knowledge they obtain, and how society may accept that knowledge.

McMillan, J. H. and Schumacher, S. (1997) *Research in Education: A Conceptual Introduction* (4th ed.). New York: Longman.

This textbook presents a comprehensive introduction to the principles, concepts and methods used in educational research. The concept maps at the beginning of each chapter serve as a good refresher, as does the list of key terms. The section on quantitative research design is excellent, and the tables found throughout the textbook are very useful.

Sapsford, R. (1999) *Survey Research*. Thousand Oaks, CA: Sage Publications.

Providing an informative coverage of the main survey methods and the place of surveys in social research, this textbook guides readers through the main theoretical and practical issues and illustrates the application of survey methods through examples. Sapsford covers the technical questions that need to be considered when planning and carrying out research. He also discusses the ethics and politics of research projects. Practical research issues are integrated with issues relating to the ethics, politics, philosophy, sociology, and social psychology that underpin them.



Tucker, M. S. and Coddling, J. B. (1998) *Standards for Our Schools: How to Set Them, Measure Them, and Reach Them*. San Francisco, CA: Jossey-Bass.

The authors offer a cogent, comprehensive argument for internationally recognized standards of achievement. The book describes the United States standards movement, explains why it is important, and outlines the steps that should be taken to achieve a standards-driven reform effort. The authors advocate not only creating a results-oriented culture devoted to continuous improvement, but also making the institution and the people in it accountable for reaching the goals set by the standards.

Vedder, P. (Ed.) (1992) *Measuring the Quality of Education*. Amsterdam: Swets and Zeitlinger.

This book contains contributions to a conference on measuring the quality of education that was organized to mark the 25th anniversary of the Institute of Education Research in 1990. While all of the chapters are relevant to those planning to undertake international studies, Chapter 10, 'IEA and the Quality of Education in Developing Countries' (T. Plomp and W. Lorly), and Chapter 11, 'Some Issues and Problems in Measuring the Quality of Education' (L. J. T. van der Kamp), seem especially relevant. Chapter 10 briefly describes IEA international comparative achievement studies, including the IEA's mission, links to national curricula, the IEA framework, and IEA target populations. Chapter 11 begins with a consideration of what educational measurement is, and then explores a range of issues including test design, study design, measurement of change, and the role of psychometrics in testing.

Vogt, W. P. (1999) *Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences* (2nd ed.). Thousand Oaks, CA: Sage Publications.

This second edition of this book covers statistical terms and methods used in social science research.

Wallance, W. A. (Ed.) (1994) *Ethics in Modeling*. Tarrytown, NY: Elsevier Science.

This book has many chapters of direct and indirect relevance to the IEA standards. For example, Chapter 5, 'Part 2: Responsible Policy Modeling', by Warren Walker, begins with these words: 'For an analyst to be credible, moral claims to truth telling must be made, for without these, why should anyone ever believe what a planner or policy analyst says.' He then goes on to describe a code of good practice similar to the central tenets of the IEA standards, including that the work should be open and explicit, and the results verifiable and reproducible. The book's main goal is to alert people to the danger of deliberately, carelessly or recklessly manipulating data to further their own specific ends.





# **Sources for Individual Standards**



# Designing, Managing and Implementing IEA Studies

## Standard for Initial Planning of an IEA Study

### Purpose

To provide guidance in specifying the design of an IEA study.

### Standard

The initial plan for an IEA study should be supplemented by a detailed study design at the earliest possible time. The design should show clearly how the study questions are to be addressed by the study, and provide a plan for the collection, processing and analysis of data, and the reporting of results.

### Summary of the Guidelines

The initial planning of an IEA study requires specifying the reasons (or justification) for the study, producing a preliminary study design, making a clear distinction between the compulsory and non-compulsory components, and providing a preliminary analysis plan, a publication and dissemination plan, a preliminary time schedule, and a preliminary international cost estimate. The initial planning standard is addressed in many of the general references listed in the preceding section, and especially in the following references.

### Sources for this Standard

Adèr Ader, H. J. and Mellenbergh, G. J. (Eds.) (1999) *Research Methodology in the Social, Behavioral and Life Sciences: Designs, Models and Methods*. Thousand Oaks, CA: Sage Publications.

This is a textbook for advanced courses in research methods and experimental design. The authors argue that the methodology of quantitative research is a unified discipline with basic notions, procedures and ways of reasoning that can be applied across the social, behavioral and life sciences. Leading contributors in their field cover the key designs, models and methods in research and seek to explain the fundamentals of the research process. The chapters on measurement models, graphical representation, experimental design and cross-sectional research are specifically relevant to many IEA studies.

Anastas, J. W. (1994) *Research Design for Social Work and the Human Services*. New York: Lexington Books.

While not specifically centered upon the educational setting, most of the concepts covered are pertinent to educational research. The section on research ethics is particularly good, as are the chapters on sampling, basic issues in data collection, and questionnaire design. The final chapter, 'Writing Research Reports', has good clear guidelines for report writing.



Boruch, R. F. (1997) *Randomized Experiments for Planning and Evaluation: A Practical Guide*. Volume 44: *Applied Social Research Methods*. Thousand Oaks, CA: Sage Publications.

After an exploration of how to judge whether or not an educational, social science or services program has worked, the author discusses the context of experiments; the standards used to judge the ethical propriety of randomized experiments; methods for studying populations; sampling methods; the randomization procedure; baseline measures; missing data registry; and analyses for quality assurance.

Campbell, D. T. (1999) *Social Experimentation*. Thousand Oaks, CA: Sage Publications.

*Social Experimentation* is a very good academic text. Parts 2 and 3, 'Threats to the Validity of Social Experiments and How They Can be Controlled' and 'Theory of Science for Social Experiments', present one of the most thorough treatments of validity in social science research.

Fink, A. (1995) *How to Design Surveys* (Vol. 5). Thousand Oaks, CA: Sage Publications.

This volume offers guidelines for quasi-experimental designs of surveys, as well as cross-sectional, cohort, and case-control designs. The book has checklists on how to avoid threats to internal and external validity.

Fink, A. (1995) *The Survey Handbook: The Survey Kit* (Vol. 1). Thousand Oaks, CA: Sage Publications.

Chapter 4, 'Reasonable Resources', identifies seven basic questions that need to be addressed when deciding what resources will be needed to conduct the survey. The chapter provides a checklist of typical survey tasks that encompass survey objects, design of the survey, preparation of the survey instrument, pilot testing, administration of the survey, organization of the data, analysis of data, and reporting.

Fink, A. (1998) *How to Conduct Surveys* (2nd ed.). Thousand Oaks, CA: Sage Publications.

This second edition of *How to Conduct Surveys* covers the following: computer-assisted and interactive surveys and how they contrast with telephone and face-to-face surveys; guidelines for preparing informed consent statements for survey respondents; ways to ensure the sample is large enough to detect a difference between groups (if one exists); ways to ask questions about ethnicity; how to read computer output containing survey results; how to prepare a structured abstract of a survey report; new survey data analysis techniques such as odds ratios, relative risks, and confidence intervals, as well as sampling techniques such as snowball sampling; and guidelines for preparing overheads and slides to report survey results, with illustrations of how an oral presentation of survey results differs from a written one.

Fowler, F. J. (1993) *Survey Research Methods*. Volume 1: *Applied Social Research Methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

*Survey Research Methods* aims to give those who collect, analyze or read about survey data a sound basis for evaluating data collection procedures. With its emphasis on the importance of minimizing non-sampling errors through good question design, good quality interviewing, and high response rates, this book will help readers understand the relationship of data collection to the reporting of figures and statistics.



Fraenkel, J. R. and Wallen, N. E. (1996) *How to Design and Evaluate Research in Education* (3rd ed.). New York: McGraw-Hill.

Chapter 5, 'Survey Research', provides an elementary introduction to survey research.

Gall, M. D., Borg, W. R. and Gall, J. P. (1996) *Educational Research: An Introduction* (6th ed.). New York: Longman.

This university textbook provides a comprehensive introduction to the major research methods and types of data analysis used in educational inquiry. Part 3, 'Planning a Research Study', includes an easy-to-read chapter on developing a research proposal.

Groves, R. M. (1989) *Survey Errors and Survey Costs*. New York: John Wiley and Sons.

The author presents a treatment of survey errors and the costs and benefits of various survey designs. More specifically, the author addresses coverage error, non-response error, sampling error, and measurement error, including the effects of interviewers and respondents, the wording of the questionnaire, and the mode of data collection. A review of the social science and statistical literatures on survey errors is included, and the relationships between the different types of errors are explored. Cost models designed to reduce various types of errors are also provided.

Krathwohl, D. R. (1998) *Methods of Educational and Social Science Research: An Integrated Approach* (2nd ed.). New York: Longman.

This textbook is also listed under the general reference section. The book's appendix, 'Writing a Research Proposal', draws upon material covered in the main part of the text, and is relevant to the development of an IEA study proposal. The appendix also has numerous tips of benefit to researchers.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

Chapter 13, 'Writing a Research Proposal', when read in association with the rest of this text, provides comprehensive guidance for this standard. This textbook is also listed under the general reference section.

Newmann, F. M. (1996) *Authentic Achievement: Restructuring Schools for Intellectual Quality*. San Francisco, CA: Jossey-Bass.

Writing from a United States perspective, the author provides a picture of the conditions under which innovations in school organization contribute to achievement. Part 1, 'Authentic Achievement', sets out standards for this type of achievement.

Orr, L. L. (1999) *Social Experiments: Evaluating Public Programs with Experimental Methods*. Newbury Park, CA: Sage Publications.

The primary focus of this textbook is social experiments. However, the treatment of many of the key concepts of social science research makes it a useful, current reference. The last chapter, 'Social Experiments and the Policy Process', has insights that are relevant to the purpose of IEA studies.



Punch, K. F. (2000) *Developing Effective Research Proposals: Essential Resource Books for Social Research*. Thousand Oaks, CA: Sage Publications.

The book is organized around three central themes. First, what are research proposals, who reads them, and why? Second, what general guidelines and strategies can help students develop a proposal? Third, what might a finished proposal look like?

Spector, P. E. (1981) *Research Designs*. Newbury Park, CA: Sage Publications.

This short but classical treatment of research design serves as an introductory textbook and a concise refresher.

Thomas, S. J. (1999) *Designing Surveys That Work: A Step-by-Step Guide*. Thousand Oaks, CA: Sage Publications.

This guide provides detailed instructions for each step of the survey process, from choosing the right topic to designing the survey, selecting and working with the respondents, and making sense of the data.



## Standard for Choosing an International Coordinating Center

### Purpose

To ensure that the international coordinating center (ICC) is properly staffed, equipped and funded to carry out its function of coordinating all aspects of study design, development, implementation, analysis and reporting.

### Standard

As the international management and co-ordination center for the study, the ICC should have access to staff with the technical and managerial competence and experience necessary to conduct the study successfully. In particular, staff will require the technical expertise to design and monitor the implementation of the study, and the managerial competence to plan and maintain realistic timelines and schedules, and to maintain close communication between the study partners, particularly national research coordinators, advisory committees and funding agencies.

The ICC should possess the infrastructure necessary to implement the study effectively, including accommodation, support staff, data processing equipment, and communications facilities (including electronic mail, fax and telephone). To maintain its operational readiness, the ICC should have access to a secure source of funds for the life of the study.

### Summary of the Guidelines

The ICC should have qualified staff, with the senior staff together having the technical knowledge and expertise to understand all aspects of the design and the operation of study, and the managerial competence and experience to oversee a large project. The center should have a sound infrastructure that includes suitable office accommodation and ready access to modern information technology, and it should also have secure funding. The references for this particular standard are derived from management and business disciplines. These references focus upon the managerial side of the research project.

### Sources for this Standard

Evans, P., Doz, Y. and Laurent, A. (Eds.) (1989) *Human Resource Management in International Firms: Change, Globalization, Innovation*. London: Macmillan.

The sections on organization culture and organizational change are relevant to this standard.



Lientz, B. P. and Rea, K. P. (1998) *Project Management for the 21st Century* (2nd ed.). San Diego, CA: Academic Press.

*Project Management for the 21st Century* focuses on proven methods and lessons learned for dealing with projects in many different industries and settings. The sections on collaborative and team project management could serve as a guide for the beginning manager.

Muchinsky, P. (2000) *Psychology Applied to Work: An Introduction to Industrial and Organizational Psychology* (6th ed.). Pacific Grove, CA: Brookes/Cole.

This graduate-level textbook has clearly stated objectives at the start of each chapter, and includes real examples throughout. The sections on personnel psychology, organizational psychology, and the work environment are all relevant to this standard.



## Standard for Formulating and Refining Study Questions

**Purpose** To ensure that the study questions are relevant to participants' needs, well chosen, well stated and can be empirically answered.

**Standard** Study questions should be clearly defined, articulated and reviewed by a wide audience to ensure that they address the critical aspects of the issues under investigation.

**Summary of the Guidelines** An international study requires the formulation of study questions that are relevant to participants' needs, are well chosen and well stated, and that can be empirically answered. The formulation process requires consultation with policy-makers, educational practitioners and funding agencies. The formulation of study questions should be devoid of false assumptions, and the questions should be written in a manner that allows them to be answered empirically. The questions also should be capable of being operationalized in terms of variables that can be validly and reliably measured, and that have the same meaning for each person.

This standard is addressed in most of the general references.

**Sources for this Standard** Bouma, G. D. (1993) *The Research Process* (rev. ed.). Melbourne, Vic: Oxford University Press.

This textbook is aimed at university students studying introductory research methods. The book assumes little knowledge of the research process. The first phase of research, as described by the author, includes selecting a research problem, narrowing the focus of the question, selecting a research design, defining and measuring variables, constructing an investigation instrument, and drawing a sample. It is this phase of research to which the author devotes the most time. The other two phases, data collection and analysis interpretation, are also addressed, but in less detail. The book is easy to read, and there are many examples that enable the reader to develop his or her understanding of the research process.

Charles, C. M. (1998) *Introduction to Educational Research* (3rd ed.). New York: Longman.

This is an introductory textbook for students interested in understanding, conducting, and interpreting both qualitative and quantitative paradigms in educational research methods. The book focuses on eight types of quantitative and qualitative research, and includes selections of non-technical, reprinted research articles to orient students who wish to conduct similar studies. Part 2, 'Preliminary Skills Needed for Conducting Research', includes an easy-to-read chapter on selecting, refining and proposing a topic for research.



Glymour, C. (1997) *Thinking Things Through: An Introduction to Philosophical Issues and Achievements*. Cambridge, MA: MIT Press.

*Thinking Things Through* provides a broad, historical and rigorous introduction to the logical tradition in philosophy and to its contemporary significance. The book focuses on three of the most fruitful issues in Western thought: What are proofs and why do they provide knowledge? How can experience be used to gain knowledge or to alter beliefs in a rational way? What is the nature of the mind and of mental events and mental states? Although this book does not have an immediate practical orientation, it serves to explore the deeper issues behind research.

Krathwohl, D. R. (1998) *Methods of Educational and Social Science Research: An Integrated Approach* (2nd ed.). New York: Longman.

The author explores the criteria of a good problem, and presents problem-development methods that biographies and literature suggest have been successful. Each section ends with either a summary of the key points or a short checklist of self-check questions. Chapter 5, 'Problem Finding', is especially relevant to this standard.

Sapsford, R. and Jupp, V. (Eds.) (1996) *Data Collection and Analysis*. London: Sage Publications.

This graduate-level textbook has two major aims: (i) to develop the ability to read research reports critically; and (ii) to develop abilities important in the production of research ideas and, more specifically, research proposals. Although the authors do describe a number of data collection and analysis methods, the book is especially relevant to the IEA standards concerned with research proposal and report writing.





## Standard for Designing an IEA Study

### Purpose

To provide guidance in specifying the design of an IEA study.

### Standard

The initial plan for an IEA study should be supplemented by a detailed study design at the earliest possible time. The design should show clearly how the study questions are to be addressed by the study, and provide a plan for the collection, processing and analysis of data, and the reporting of results.

### Summary of the Guidelines

An IEA study requires careful statement of the study questions. The next step after this is to determine the methodological approach of the study and to provide a description of how the study will address the study questions. This necessitates a detailed sampling design that will provide precise and economical estimates of population parameters and the identification of data collection instruments and of data collection and analysis methods. Many of the references identified for this standard draw either from quantitative research methodology, survey design, and/or sampling theory. Many of the general references pertaining to research methods cover this standard.

### Sources for this Standard

Charles, C. M. (1998) *Introduction to Educational Research* (3rd ed.). New York: Longman.

This is an introductory textbook for students interested in understanding, conducting and interpreting both qualitative and quantitative paradigms in educational research methods. Part 3, 'Conducting Your Own Research Project', includes a chapter on designing a research project.

Kish, L. (1987) *Statistical Design for Research*. New York: John Wiley and Sons.

This book addresses basic aspects of research design that are central and common to many related fields in the social sciences, the health sciences, education, and market research. Kish presents a unified approach to a common core of problems of statistical design that exists in all these fields, along with basic similarities in practical solutions. He takes a broader, more general and philosophical view of the statistics for the more fundamental aspects of design than do the standard treatments of experimental design. This book can serve as both a textbook and a reference manual.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

This book is designed for people new to research and planning who intend to undertake a research project. Chapter 4, 'Formulating a Research Problem', Chapter 7, 'The Research Design', and Chapter 8, 'Selecting a Study Design', are especially relevant to this standard.



McMillan, J. H. and Schumacher, S. (1997) *Research in Education: A Conceptual Introduction* (4th ed.). New York: Longman.

The material of particular relevance to the standards is found in Chapter 16, 'Guidelines for Research Proposals', which has an excellent presentation of the research proposal process. Three sections of the chapter deserve particular attention: 'Forms of Research Communication and Reasoning', 'Quantitative Research Proposals' and 'Preparation and Criticism of a Proposal'.

Rosier, M. J. (1997) Survey research methods. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 154–61). Cambridge, UK: Pergamon.

This article describes the purposes of survey research and the steps to take when conducting survey research. The steps form a cycle: (a) formulation of the research questions, (b) definition of a conceptual framework, (c) development of survey instruments, (d) collection of data, (e) preparation of the data analysis, (f) conduct of analysis, and (g) reporting of the findings of the survey.

Thurlow, M. L., Elliott, J. L. and Ysseldyke, J. E. (1998) *Testing Students with Disabilities: Practical Strategies for Complying with District and State Requirements*. Thousand Oaks, CA: Corwin Press.

This book is written from a United States perspective to provide practitioners with information about testing students with disabilities. The authors provide a range of practical strategies for accomplishing this goal. Chapter 1 presents reasons why students with disabilities should be included in testing. These reasons include the need to present an accurate picture of education, and to avoid unintended consequences of exclusion. Chapter 3, 'Assessment Accommodation: Who is Eligible?', and Chapter 4, 'Accommodations to Consider', present a comprehensive picture of the considerations needed to test students with disabilities. The book concludes with a number of forms, including 'Participation Decision Making', 'IEP Form for Identifying Accommodations with Examples', 'Test Accommodation Planning Chart' and 'Logistic and Strategic Plans for Providing Assessment Accommodations'.



## Standard for Developing a Sampling Plan

### Purpose

To ensure that the data provided by study sample(s) from each country represent the international target population(s) with a level of accuracy that allows the study questions to be answered.

### Standard

The sampling plan should ensure that national target population(s) are defined operationally in such a way that they can be represented accurately using sample survey methodology. The sampling design should provide a detailed plan for estimating the parameters of the target population(s) with samples of the smallest possible size and greatest possible cost-effectiveness, while maintaining an acceptable level of accuracy.

### Summary of the Guidelines

The development of an IEA sampling plan begins with the specification of the international population. The target population is then operationally defined so as to be amenable to effective probability sampling. At the country level, the 'national defined population' is the operational definition of the 'national desired population'. A description of the sampling plan, and the required sampling precision and the minimum sample size for each country all form part of the sampling plan. The sampling design should represent the best compromise within the range of tolerable sampling errors, costs, burden and other considerations.

### Sources for this Standard

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A. and Sudman, S. (1991) *Measurement Errors in Surveys*. New York: John Wiley and Sons.

This book presents emerging principles and trends in measurement error in surveys. It reports new research findings and promotes interdisciplinary exchanges in numerous approaches to the assessment, modeling and reduction of measurement inaccuracies in surveys. The book features sections on the questionnaire and other means of data collection, the respondent-interviewer relationship, and the effects of measurement errors on estimation and data analysis.

Braverman, M. T. (1996) Sources of survey error: implications for evaluation studies. In M. T. Braverman and J. K. Slater (Eds.), *Advances in Survey Research* (pp. 17–28). San Francisco, CA: Jossey-Bass.

The author briefly overviews some recent advances in survey research within the context of different kinds of survey error.



Foy, P. and Joncas, M. (2000) TIMSS sample design. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 29–48). Chestnut Hill, MA: Boston College.

This chapter outlines the TIMSS 1999 sampling design.

Foy, P., Rust, K. and Schleicher, A. (1996) Sample design. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study: Technical Report. Volume 1: Design and Development* (pp. 4-1–4-5). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1995 technical report describes pertinent aspects of the TIMSS 1995 sampling plan. More specifically, it describes the forms used to document sampling procedures, the sample design, sample execution, population definitions and sample participation rates, and the reporting of achievement results.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1993) *Sample Survey Methods and Theory. Volumes 1 and 2: Methods and Applications*. New York: John Wiley and Sons.

This general-purpose work covers sampling method and theory. Volume 1 gives a simple, non-mathematical discussion of principles and their practical applications. Volume 2 covers theory and proofs.

Henry, G. T. (1990) *Practical Sampling* (Applied Social Research Methods Series, Vol. 21). Newbury Park, CA: Sage Publications.

This short book provides a clear treatment of sampling theory, logic of design, and execution of the design. The central aim of the practical sampling design, to improve the validity of the survey results, is reinforced.

Jaeger, R. M. (1985) *Sampling in Education and the Social Sciences*. New York: Longman.

The first chapter provides an excellent overview of survey research. Chapter 2, 'Basic Concepts of Survey Theory', Chapter 4, 'Stratified Samples', and Chapter 7, 'Single Stage Cluster Sampling: Procedures, Estimation Formulas and Application', seem particularly relevant to the design of IEA studies.

Kalton, G. (1983) *Introduction to Survey Sampling*. Beverly Hills, CA: Sage Publications.

This book is primarily concerned with sample design issues. The author leads the reader through the concepts of sample design, and the sections on weights, stratification, and cluster design are particularly good.

Kish, L. (1995) *Survey Sampling*. New York: John Wiley and Sons.

This accessible book on sampling techniques emphasizes surveys of human populations and draws illustrations from them. Kish explains how to design and execute valid samples of moderate dimensions and difficulty, how to avoid selection biases, and how to become more adept at evaluating sample results by judging their validity and limits of inference, applicability and precision.



Krathwohl, D. R. (1998) *Methods of Educational and Social Science Research: An Integrated Approach* (2nd ed.). New York: Longman.

Chapter 8 of this master's-level textbook covers the principles of sampling, probability and non-probability sampling, probability sampling procedures, generality, external validity, and population definition. The chapter has a series of excellent summary tables. For example, one table has a description of the advantages and disadvantages of random and stratified sampling.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

This practical book is designed for people new to research and planning who intend to undertake a research project. Chapter 12, 'Sampling', gives an easy-to-read, comprehensive coverage of sampling methodology.

Lehtonen, R. and Pahkinen, E. J. (1995) *Practical Methods for Design and Analysis of Complex Surveys*. New York: John Wiley and Sons.

This book provides readers with practical, up-to-date tools for basic and more advanced sampling and estimation, and demonstrates the benefits of using available auxiliary information for efficient sampling and estimation. The authors describe approximative techniques for variance and covariance estimation as well as analysis methods for frequency tables and multivariate techniques usable in complex surveys. Each method is illustrated with numerical examples and case studies from actual surveys.

Levy, P. S. and Lemeshow, S. (1999) *Sampling of Populations: Methods and Applications* (3rd ed.). New York: John Wiley and Sons.

This sampling textbook provides a highly readable, practical treatment of the subject. Keeping the mathematics to a minimum, it walks the reader through real-world sample surveys—from sampling designs to problems of missing data and non-response to estimation procedures. It also includes multistage sampling, analysis of sample survey data, and software manipulation.

Maisel, R. and Persell, C. H. (1995) *How Sampling Works* (Pine Forge Press Series in Research Methods and Statistics). Thousand Oaks, CA: Sage Publications.

The principles of statistical sampling and inference are explained in this hands-on volume. Software packaged with the book helps to clarify these concepts for professionals and students in a non-mathematical way.

McMillan, J. H. and Schumacher, S. (1997) *Research in Education: A Conceptual Introduction* (4th ed.). New York: Longman.

The chapter in this book relevant to the standard is Chapter 5, a portion of which is devoted to sampling. The authors discuss the strengths and weaknesses of sampling methods and summarize them in a table.



Namboodiri, N. K. (Ed.) (1978) *Survey Sampling and Measurement*. New York: Academic Press.

This reference book offers an excellent introductory coverage of the topics in survey research.

Ross, K. N. (1988) Sampling. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (pp. 527–37). Oxford, UK: Pergamon.

This article offers an excellent introduction to practical sampling in a school setting.



## Standard for Choosing Data Collection Methods

### Purpose

To ensure the selection of the most appropriate and effective methods for collecting data to address the study questions.

### Standard

The data collection methods should be chosen so as to provide the information necessary to address the study questions in a manner that minimizes the burden on the data providers (that is, questionnaire respondents, test-takers, etc.) and makes the least demand on available resources.

### Summary of the Guidelines

The most appropriate and effective methods for collecting data to address the study questions should be chosen. The methods should place the least burden on the respondents, and utilize available resources. This standard requires the specification of the sources of data, the collection (including timing), processing, analysis and reporting procedures to be used, and a description of the inferences to be made from the data.

### Sources for this Standard

Krosnick, J. A., Narayan, S. and Smith, W. R. (1996) Satisficing in surveys: initial evidence. In M. T. Braverman and J. K. Slater (Eds.), *Advances in Survey Research* (pp. 29–44). San Francisco, CA: Jossey-Bass.

This article examines the cognitive burden that a respondent experiences, the consequences of such a burden, and the methods that can be taken to prevent or minimize it.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

This practical book is designed for people new to research and planning who intend to undertake a research project. Chapter 9, 'Selecting a Method of Data Collection', is directly relevant to this standard.

Stouthamer-Loeber, M. and van Kammen, W. B. (1995) *Data Collection and Management: A Practical Guide* (Applied Social Research Methods Series, Vol. 39). Thousand Oaks, CA: Sage Publications.

The book emphasizes the following: applying good standards for data collection and processing; planning research activities; making budgetary decisions and exercising budgetary control; hiring, training and supervising field interviewing staff; and ensuring high participant acquisition and retention rates. The last item includes techniques for locating lost participants.



## Standard for Developing a Quality Assurance Program

### Purpose

To ensure that all aspects of the study are conducted to a high standard, and that the data collected are of sufficiently high quality to support authoritative international comparisons. Although quality control should be a feature of all work at both international and national centers, it is particularly important for activities such as test administration, which may be conducted by school personnel and therefore be outside the control of study staff.

### Standard

The international coordinating center (ICC) should ensure that all operational documentation emphasizes quality control as an integral part of every activity. In particular, data collection activities should include provision for independent monitoring of a sample of data collection sites. Such independent monitoring should be conducted by the national center and also, as far as funds permit, by the international coordinating center (see also **Standard for Implementing Data Collection Quality Control Procedures**).

### Summary of the Guidelines

The international coordinating center should ensure that all operational documentation emphasizes quality control as an integral part of every activity. This requires the development of procedures for checking and re-checking materials, and making explicit provision for resources for quality assurance. The verification of the accuracy of translations, the use of quality control observers for monitoring data collection activities, and the independent checking of analyses all form part of the quality assurance program. More broadly, issues of quality control in education research can draw upon lessons learned in business.

### Sources for this Standard

Boruch, R. F. (1997) *Randomized Experiments for Planning and Evaluation: A Practical Guide* (Applied Social Research Methods Series, Vol. 44). Thousand Oaks, CA: Sage Publications.

The author discusses, amongst other concepts, analyses for quality assurance.

Flood, R. L. (1993) *Beyond TQM*. New York: John Wiley.

Flood presents a complete analysis regarding the strengths and weaknesses of total quality management (TQM) and explores the opportunities and misconceptions that come with it. The author also provides a diverse set of case studies that show TQM in action.





Martin, M. O., Mullis, I. V. S. and Kelly, D. L. (1996) Quality assurance procedures. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development* (pp. 11-1–11-12). Chestnut Hill, MA: TIMSS International Study Center.

This chapter from the TIMSS 1995 technical report describes the procedures used to ensure high-quality data across all countries, and the support afforded to the national centers by the international study center in the form of standardized manuals, software aids, practical training and technical assistance.

O'Connor, K. M. and Stemler, S. E. (2000) Quality control in TIMSS data collection. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 135–56). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report summarizes quality assurance activities for the TIMSS 1999 data collection.

Sinha, M. N. and Willborn, W. W. O. (1985) *The Management of Quality Assurance*. New York: John Wiley.

Sinha and Willborn summarize the essential concepts, practices and methods of modern quality assurance management, balancing statistical techniques with a managerial approach.

Wadsworth, H. M., Stephens, K. S. and Godfrey, A. B. (1986) *Modern Methods for Quality Control and Improvement*. New York: John Wiley.

In this comprehensive presentation of modern quality control methods and systems, the authors explore the traditional principles and techniques of statistical quality control and quality assurance. The book is designed for advanced undergraduate or graduate students in industrial engineering and management.



## Standard for Preparing an Analysis Plan

### Purpose

To provide a detailed plan that clearly shows how the study data will be analyzed to achieve the study objectives. The plan should enable NRCs and other reviewers to evaluate the intended analytic techniques and statistical procedures in the light of the study's aims, scope and resources.

### Standard

The analysis plan should list each of the study's goals or research questions, and provide details of the techniques to be applied to address the issues, together with a justification for each technique. Sufficient detail should be provided to enable reviewers to evaluate the utility and suitability of the application.

### Summary of the Guidelines

The analysis plan addresses the stated aims of the study and specifies the analysis to be performed. The analyses should select the most appropriate statistical technique and make clear how the analytic techniques chosen match the design of the study. Some of the references presented below are specifically aimed at developing the analysis plan, while others refer to analysis techniques. Many of the general references listed at the start of this bibliography also address the preparation of an analysis plan.

### Sources for this Standard

Fink, A. (1995) *How to Analyze Survey Data* (Vol. 8). Thousand Oaks, CA: Sage Publications.

This volume shows what statistics can do for surveys, describes Steven's measurement scales in detail, and demonstrates how to choose a method to analyze survey results. The book includes checklists, guidelines and practice exercises. The volume also discusses regression and hypotheses testing, and offers a cogent discussion on screening and transforming data.

Green, K. E. (1996) Application of the Rasch model to evaluation of survey data quality. In M. T. Braverman and J. K. Slater (Eds.), *Advances in Survey Research* (pp. 81–92). San Francisco, CA: Jossey-Bass.

Green provides an easy-to-read overview of validity and the Rasch model. The basic application of the Rasch model to data and the item-person map are particularly clear.

Hoyle, R. H. (Ed.) (1995) *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications.

This collection of articles represents a comprehensive synthesis of structural equation modeling (SEM). Chapter 1 presents the basic concepts of SEM, and subsequent chapters deal with a diverse range of issues that include model specification, evaluating model fit, and statistical power in SEM. This book is for advanced graduates.



Kolen, M. J. and Brennan, R. L. (1999) *Test Equating: Methods and Practices* (Springer Series in Statistics). New York: Springer-Verlag.

This textbook and reference book provides a practically oriented introduction to test equating that discusses the most frequently used equating methodologies and covers many of the practical issues involved. The main themes are: the purpose of equating; distinguishing between equating and related methodologies; the importance of test equating to test development and quality control; the differences between equating properties, equating designs, and equating methods; equating error; and the underlying statistical assumptions for equating.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

This practical book is designed for people new to research and planning who intend to undertake a research project. Chapter 15, 'Processing,' guides the reader through the major considerations in processing the data.

Lee, E. S., Forthofer, R. N. and Lorimor, R. J. (1989) *Analyzing Complex Survey Data* (Quantitative Applications in the Social Sciences, Vol. 71). Thousand Oaks, CA: Sage Publications.

*Analyzing Complex Survey Data* offers not only a means of examining and analyzing complex surveys, but also guidance on how to overcome the problems that often arise. It includes discussions on the choices involved in variance estimates, simple random sampling without replacement, stratified random sampling and two-stage cluster sampling, and outlines the different computer programs that are currently available.

Lewis-Beck, M. S. (1995) *Data Analysis: An Introduction* (Quantitative Applications in the Social Sciences, Vol. 103). Thousand Oaks, CA: Sage Publications.

This introduction to data analysis focuses on the interpretation of statistical results, in particular those that come from non-experimental social research. It provides social science researchers with the tools necessary to select and evaluate statistical tests appropriate for their research question. Using a consistent data set throughout the book to illustrate the various analytical techniques, Michael Lewis-Beck covers such topics as univariate statistics, measures of association, the statistical significance of the relationship between two variables, simple regression in which the dependent variable is influenced by a single independent variable, and multiple regression.

Lindsey, J. K. (1999) *Applying Generalized Linear Models*. New York: Springer-Verlag.

The author shows the unity of many of the commonly used models and provides readers with a taste of many different areas, such as survival models, time series and spatial analysis, and of their unity. The reader is assumed to have knowledge of basic statistical principles, whether from a Bayesian, frequentist, or direct likelihood point of view, and to be familiar with at least the analysis of the simpler normal linear models, regression and ANOVA.



Little, R. J. A. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.

The authors review historical approaches to the subject, and describe rigorous yet simple methods for multivariate analysis with missing values. They then go on to provide a coherent theory for analysis of problems based on likelihoods derived from statistical models for the data and the missing data mechanism. The theory is applied to a wide range of important missing-data problems. This textbook is for graduates studying at an advanced level.

Tacq, J. (1997) *Multivariate Analysis Techniques in Social Science Research: From Problem to Analysis*. London: Sage Publications. (Originally published in Dutch as *Van Probleem naar Analyse* by Rotterdams Instituut voor Sociologisch en Bestuurskundig Onderzoek, Rotterdam, 1991.)

The development of an analysis plan requires the linking of the research problem to the techniques of analysis. Using a number of research examples, and laying out the research design in clear steps, Tacq shows how this linkage can be achieved. Chapter 3, 'The Analysis Techniques as the Mirror of the Research Problem', is excellent. In this chapter, Tarq leads the reader through the process of recognizing the structure of the research problem, the formation of an appropriate research design, and the anticipation of a suitable analysis technique.

van der Linden, W. J. and Hambleton, R. K. (Eds.) (1997) *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.

This book represents one of the most comprehensive presentations of item response theory (IRT). The 28 chapters provide detailed accounts of models, applications and issues in IRT.



# Developing Data Collection Instruments

## Standard for Developing Assessment Frameworks and Conceptual Models

### Purpose

To provide a structure for the study that ensures the instruments developed address the research questions, reflect recent curricular emphases and learning objectives across the participating countries and include what various scholars, practitioners and interested citizens believe should be in the assessment. If necessary, the instruments should also maintain ties to previous assessments to permit the reporting of trends in student achievement, backgrounds and experiences across time.

### Standard

The framework or conceptual model underlying instrument development for the study should define the subject area domain that will be assessed in the study (for example, mathematics, science, civics, etc.). It should:

- (i) Describe the underlying constructs and procedures followed for defining the domain to be assessed.
- (ii) Specify the dimensions, areas, topics and subtopics within the domain that are age/grade appropriate for the target population.
- (iii) Explain the purpose for which the results from the tests and questionnaires should be used.
- (iv) Provide structure for describing what students should know and be able to do in the subject area being assessed. This structure will provide the basis for developing specifications for developing cognitive items and questionnaires and, eventually, for reporting the study results.

### Summary of the Guidelines

The conceptual model provides structure for the study, and ensures the instruments developed address the research questions, reflect current curricula emphases and learning objectives, and include what leading experts believe should be assessed. The conceptual model should be applicable to the various participating countries, reflect experience gained from previous assessments and research efforts, and take into account national reform efforts. The references listed below range from technical books (how to develop an assessment framework) through to more general ones that present alternative assessment ideas.



## Sources for this Standard

Airasian, P. W. (1994) *Classroom Assessment*. New York: McGraw-Hill.

While this university textbook gives an excellent presentation of classroom assessment, Chapters 1, 5, 6 and 7 are especially pertinent to IEA-type studies. Airasian covers all aspects of testing in a clear manner, supporting key points with examples. He uses tables to provide excellent summaries of key points.

Airasian, P. W. (1996) *Assessment in the Classroom*. New York: McGraw-Hill.

This well-written textbook is aimed at the school practitioner. Chapter 3, 'Formal Assessment: Teacher-Made Tests', has a very good presentation of the considerations of test-making and test administration. The summary tables are especially useful. The section, 'General Guidelines for Writing and Critiquing Test Items', is pertinent to this IEA standard. The textbook as a whole serves as a useful introduction to the notion of school assessment.

Berlak, H., Newmann, F. M., Adams, E., Archibald, D. A., Burgess, T., Raven, J. and Romberg, T. A. (1992) *Toward a New Science of Educational Testing and Assessment*. Albany, NY: State University of New York Press.

This book presents arguments for authentic assessment. Of particular interest is Berlak's Chapter 8, in which he questions psychometricians who claim to provide scientifically valid means of measuring academic achievement. Berlak calls for a transition away from the psychometric paradigm of standardized testing toward a system of contextual assessment. The book is relevant to the IEA standards in as much as it presents cogent arguments about assessment and challenges the psychometric paradigm.

Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O. and Sainsbury, M. (2001) *Framework and Specifications for PIRLS Assessment 2001* (2nd ed.). Chestnut Hill, MA: Boston College.

This publication presents the assessment framework and specifications for IEA's Progress in International Reading Literacy Study (PIRLS) for 2001. It contains a good example of an assessment framework for reading literacy for fourth-grade students (9 to 10 year olds), and of the conceptual framework for studying the home, school, community and national contexts within which students learn to read. It also outlines the assessment design and the specifications for achievement tests and questionnaires, and provides examples of passages and items used in the PIRLS assessment.

De Landsheere, V. (1997) Taxonomies of educational objectives. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 803–12). Cambridge, UK: Pergamon.

This article provides a brief presentation of cognitive, affective and psychomotor taxonomies. The cognitive taxonomies presented are Bloom's taxonomy, Guilford's structure of intellect model, the Gagné-Merrill taxonomy, and De Block's taxonomy.



Frederiksen, N., Mislevy, R. J. and Bejar, I. I. (1993) *Test Theory for a New Generation of Tests*. Hillsdale, NJ: Lawrence Erlbaum.

This collection of articles presents arguments for the development of a new psychometric test theory and describes the limitations of the classical test theory. The proposed new theory evolves from item response theory, but links with a broader understanding of the learner. This book is written for an academic audience, and argues that standard test theory is missing 'just how people know what they know, what they can do and the ways in which they can increase their capacities' (p. ix).

Gronlund, N. E. (1998) *Assessment of Student Achievement* (6th ed.). Boston, MA: Allyn and Bacon.

The first three chapters of this textbook pertain specifically to this standard. Chapters 1 and 2 set the broad framework of assessment, while Chapter 3, 'Planning the Achievement Test', includes sections of direct, practical relevance to the standard. The section, 'Preparing the Test Specifications' (pp. 37–42), explains why specifications are required and how they can be utilized in constructing the test. The preparation of a two-way table of specifications is described. This table relates the outcomes to course content and indicates the relative weight that each outcome is to be given.

Henning-Stout, M. (1994) *Response Assessment: A New Way of Thinking about Learning*. San Francisco, CA: Jossey-Bass.

Starting with the nature of learning in schools, Henning-Stout examines the ways academic assessment can be used responsively to support learning and provide students with a sense of themselves as learners. The engagement of learners in their own assessment is supported by case examples. Part 2, 'Guidelines for Practice', includes chapters on assessment and intervention in the learning environment, and mathematical reasoning.

Hoi, K. and Suen, H. K. (1990) *Principles of Test Theory*. Mahwah, NJ: Lawrence Erlbaum.

This book offers a summary of core issues in test theory. In so doing, it provides a comprehensive survey of reliability, validity and item analysis from the perspectives of classical true-score model, generalizability theory, item response theory, criterion-referenced testing and behavioral assessment. The authors also discuss related theoretical issues, such as item bias, equating and cut-score determination.

Keeves, J. P. (1997) Validity of tests. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 976–84). Cambridge, UK: Pergamon.

This article considers tests and their interrelationship with the curriculum and with instruction. Also included are several indicators that can be employed to show the extent to which a test may be considered to be a valid instrument for measuring learning.



Kulm, G. (1994) *Mathematics Assessment: What Works in the Classroom*. San Francisco, CA: Jossey-Bass.

The author offers hands-on guidelines for assessing student understanding of procedures, concepts and problem-solving. More specifically, the guidelines include advice on developing and interpreting assessment tasks, assessing individual and group mathematics activities, using scoring rubrics, and developing student portfolios. The focus is upon what works within the classroom, and the book covers a wide range of alternative assessment techniques. Part 1, 'Background and Perspectives', covers the purposes and goals of assessment, assessment and mathematics teaching, procedural and conceptual knowledge, and problem solving and strategic knowledge.

Millman, J. and Greene, J. (1989) The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 335–66). New York: American Council on Education and Macmillan Publishing.

This chapter (Chapter 8) is directed to the professional test constructor, and emphasizes options for specifying and developing tests. The discussion begins with the purpose of the test, possible elements of a test plan, the types of items and how they are written, the evaluation and selection of items, and a brief section on putting the items together into a test.

Mullis, I.V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J. and Chrostowski, S. J. (2001) *TIMSS Assessment Frameworks and Specifications 2003*. Chestnut Hill, MA: Boston College.

This publication presents the TIMSS assessment framework and specifications for the 2003 and subsequent assessments. It contains good examples of mathematics and science assessment frameworks for fourth- and eighth-grade students, and of the contextual framework for studying the curriculum and home, school and community contexts within which students learn mathematics and science. It also provides details of the matrix-sampling assessment design and the specifications for achievement tests and questionnaires.

Newmann, F. M. (1996) *Authentic Achievement: Restructuring Schools for Intellectual Quality*. San Francisco, CA: Jossey-Bass.

Writing from a United States perspective, the author provides a picture of the conditions under which innovations in school organization can contribute to achievement. Part 1, 'Authentic Achievement', sets out standards for this type of achievement.

Ory, J. C. and Ryan, K. E. (1993) *Tips for Improving Testing and Grading* (Survival Skills for Scholars). Newbury Park, CA: Sage Publications.

This introductory textbook is aimed at the school-practitioner. Chapter 1 commences with the need for course objectives, briefly describes assessing different levels of learning (for example, Bloom's taxonomy), and concludes with developing a testing blueprint. Chapter 2 presents and then answers seven guiding questions for developing a test: What is the purpose of the exam? How difficult should you make the exam? Who is taking the exam? How many students are taking the test? How much time has been provided for testing? How many exams are scheduled? What type of test is desirable?





Phye, G. D. (Ed.) (1997) *Handbook of Classroom Assessment: Learning, Achievement, and Motivation*. San Diego, CA: Academic Press.

The *Handbook of Classroom Assessment* takes a multi-dimensional approach to classroom assessment. Most chapters are devoted to the assessment of learning and achievement, and discuss current theories. The book also features assessment of academic self-concept and subjective wellbeing in children and adolescents. The handbook provides successfully field-tested examples of assessment techniques and strategies within the content areas of mathematics, social studies, foreign languages and the visual arts.

Rothman, R. (1995) *Measuring Up: Standards, Assessment, and School Reform*. San Francisco, CA: Jossey-Bass.

Writing from a United States perspective, Rothman argues that school reform and assessment must be driven by what students are expected to know and be able to do.

Tucker, M. S. and Coddling, J. B. (1998) *Standards for Our Schools: How to Set Them, Measure Them, and Reach Them*. San Francisco, CA: Jossey-Bass.

The authors offer a cogent, comprehensive argument for internationally recognized standards of achievement. The book describes the United States standards movement, explains why it is important, and what steps should be taken to achieve a standards-driven reform effort. The authors advocate creating a results-oriented culture devoted to continuous improvement, and making the institution and the people in it accountable for reaching the goals set by the standards.

Wiggins, G. P. (1998) *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco, CA: Jossey-Bass.

This book provides guidance on how to design performance-based assessment. The book leads its primary audience, teachers, through aspects of assessment design, including how to design performance tasks to meet educational standards, how to score assessments fairly, and how to structure and judge student portfolios. Wiggins also looks at how performance assessment can be used to improve curriculum and instruction, grading and reporting, and teacher accountability. Part 1, 'Essential Elements of Assessment', provides a cogent argument for school-based assessment that aims to improve, rather than to audit, student achievement.



# Standard for Developing Specifications for Tests and Questionnaires

## Purpose

To ensure that the instruments will reflect the framework developed to assess the subject area, are consistent with the resources available for the study, and will provide data that can be analyzed to address the study's research questions. There also needs to be an assurance that the inferences drawn from these results are likely to be consistent with their intended purpose.

## Standard

The specifications for the tests and questionnaires should describe the content of the instruments, the types of items to be used, and the timing and the conditions (for example, the physical environment, reference materials, laboratory equipment, etc.) under which the instrument is administered. They should also describe how the responses will be scored.

## Summary of the Guidelines

The tests and questionnaires must reflect the conceptual model, be consistent with the resources available, and provide data that can be analyzed to address the study's research questions. These requirements necessitate the following: a description of the medium and format of the assessment instruments and response form; an explanation of the design for the instrument development; a specification of the questions, tasks and items that should be in the instruments; an estimation of the number of questions, tasks and items that should be included; and a description of any special equipment. The desired psychometric characteristics of the tests and questionnaires should be specified, as should the procedures for scoring the items.

## Sources for this Standard

Airasian, P. W. (1994) *Classroom Assessment*. New York: McGraw-Hill.

While this university textbook gives an excellent presentation of classroom assessment, Chapters 1, 5, 6 and 7 are especially pertinent to IEA-type studies. Airasian covers all aspects of testing in a clear manner, supporting key points with examples. He uses tables to provide excellent summaries of key points. For example, Table 6.02, 'Suggestions for Preparing Test Items,' encapsulates the accumulated wisdom pertinent to writing good test items.

Airasian, P. W. (1996) *Assessment in the Classroom*. New York: McGraw-Hill.

This well-written, short textbook is aimed at the school practitioner. Chapter 3, 'Formal Assessment: Teacher-Made Tests,' has a very good presentation of the considerations of test-making and test administration. The summary tables are especially useful. The section 'General Guidelines for Writing and Critiquing Test Items' is pertinent to this IEA standard.



Aitken, L. R. (1998) *Tests and Examinations: Measuring Abilities and Performance*. New York: John Wiley and Sons.

This book provides brief but complete coverage of concepts, methods and materials for selecting, designing and using tests of achievement, aptitudes and special abilities. Aitken provides step-by-step guidelines along with a set of computer programs for planning, constructing, administering, scoring and evaluating tests in educational, clinical, business/industrial and government/military settings.

Cangelosi, J. S. (1990) *Designing Tests for Evaluating Student Achievement*. New York: Longman.

Chapter 1 of this university textbook sets the framework of a test by defining the learning goals students are expected to achieve. It specifies the subject matter and behavioral construct of each goal. Throughout the chapter, the author uses cognitive, affective and psychomotor taxonomies.

Feldt, L. S. and Brennan, R. L. (1989) Reliability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 105–46). New York: American Council on Education and Macmillan Publishing.

In this chapter, the authors define the essence of reliability analysis as the quantification of the consistency and inconsistency of examinee performance. The presentation of classical test theory and of generalizability-related areas of reliability is well covered.

Fink, A. (1995) *How to Ask Survey Questions* (The Survey Kit, Vol. 2). Thousand Oaks, CA: Sage Publications.

This volume guides readers in determining a feasible number of survey questions, preparing appropriately worded questions for participants, and standardizing the response choice. It also covers such issues as how to understand a survey's cultural, psychological, economic and political context; how to ask questions about knowledge, attitudes and behaviors; how to avoid negative phrasing and biased words; and how to distinguish among response formats that use nominal, ordinal and numerical measurement.

Goldstein, H. and Lewis, T. (1996) *Assessment: Problems, Developments and Statistical Issues*. New York: John Wiley and Sons.

The book covers such areas as the different purposes of assessment, questions of design and structure, control of grading procedures, locally based assessment, and social and political aspects. Leading experts in their respective fields provide in-depth examinations of topical issues. The wider social and educational implications of different assessment procedures are reviewed.

Gronlund, N. E. (1998) *Assessment of Student Achievement* (6th ed.). Boston, MA: Allyn and Bacon.

Chapter 4, 'Writing Selection Items: Multiple-Choice', Chapter 5, 'Writing Selection Items: True-False, Matching, and Interpretive Exercise', and Chapter 6, 'Writing Supply Items: Short Answer and Essay', each describe how test items should be written. Gronlund provides clear, practical rules for the writing of each item type within the framework provided by the test specification. While some theoretical considerations are presented in the book, the main focus is on enabling people (for example, teachers) to develop good items.



Haladyna, T. M. (1999) *Developing and Validating Multiple Choice Test Items* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

This book addresses the related topics of multiple-choice test item development and validation of responses to these test items. In so doing, the volume provides a conceptual basis for item writing, reviews the issue of constructed-response versus selected-response testing, presents a variety of formats, and gives guidance in developing items. It also provides a basis for reviewing, evaluating and improving items, and speculates about the future of item development and validation.

Krathwohl, D. R. (1998) *Methods of Educational and Social Science Research: An Integrated Approach* (2nd ed.). New York: Longman.

The author explores the criteria of a good problem and presents problem development methods that biographies and literature suggest have been successful. Chapter 16, 'Survey Research: Questionnaires', provides an excellent analysis of the place of questionnaires in research, and makes many points pertinent to good survey and questionnaire design.

Kulm, G. (1994) *Mathematics Assessment: What Works in the Classroom*. San Francisco, CA: Jossey-Bass.

Kulm offers hands-on guidelines for assessing student understanding of procedures, concepts and problem-solving. More specifically, the guidelines include advice on developing and interpreting assessment tasks, assessing individual and group mathematics activities, using scoring rubrics, and developing student portfolios. The focus is on what works in the classroom, and the book covers a wide range of alternative assessment techniques. Part 2, 'Planning and Designing an Assessment Program', offers many practical suggestions for mathematics assessment. The final chapter in this part, Chapter 10, 'Scoring and Grading Techniques', is especially relevant to this standard.

Messick, S. (1989) Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 13–103). New York: American Council on Education and Macmillan Publishing Company.

This article (Chapter 1) is one of the most comprehensive presentations of validity ever given. Messick amplifies two basic concepts: (i) validity is unified through the facet concept; and (ii) validation is scientific inquiry.

Mintzes, J. J., Wandersee, J. H. and Novak, J. D. (Eds.) (1998) *Teaching Science for Understanding*. San Diego, CA: Academic Press.

*Teaching Science for Understanding* is written from a human constructivist perspective. It begins with an overview of the changes in science education, and then presents a review of each major instructional strategy, information about how each strategy is best used, and the effectiveness of the strategies for understanding and retaining information. The book presents the main strategies used to achieve this depth of understanding, including the use of computer simulations, small laboratories, and journal writing, and discusses how to use each strategy at the elementary, secondary and university levels.



Mintzes, J. J., Wandersee, J. H. and Novak, J. D. (Eds.) (2000) *Assessing Science Understanding*. San Diego, CA: Academic Press.

*Assessing Science Understanding*, like its companion volume *Teaching Science for Understanding*, is written from a human constructivist perspective. In exploring how to assess whether learning has taken place, the book discusses a range of promising new and practical tools for assessment that include concept maps, vee diagrams, clinical interviews, problem sets, performance-based assessments, computer-based methods, visual and observational testing, portfolios, explanatory models and national examinations.

Spector, P. E. (1992) *Summated Rating Scale Construction: An Introduction* (Quantitative Applications in the Social Sciences, Vol. 82). Thousand Oaks, CA: Sage Publications.

Aimed at helping researchers construct more effective scales, this book shows readers how to determine the number of items necessary, the appropriate amount of response categories, and the most productive wording of items. The author also provides guidance on how to sort good items from bad, and how to validate a scale, including dimensional validity from factor analysis. The book concludes with a step-by-step account of how to develop a summated rating scale based on classical test theory.

Thorndike, R. L. and Thorndike, R. M. (1997) Reliability. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 775–90). Cambridge, UK: Pergamon.

This article, in seeking to answer the question of how accurately the observation sample represents the broader responses from which it was drawn, provides a thorough coverage of reliability.

Zeller, R. A. (1997) Validity. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 822–29). Cambridge, UK: Pergamon

The author defines and contrasts validity with reliability. Zeller also describes the various types of validity and proposes a new definition of validity.



## Standard for Test Development

### Purpose

To ensure that the tests developed by IEA provide fair and accurate measures of students' achievement on the subject matter domain defined by the framework, and that they adhere to the test specifications.

### Standard

The tests as a whole should meet domain definitions and test specifications. All aspects of the test need to be clear and accurate, including the directions, stimulus materials and items. The individual items and tasks must be appropriate to the purpose of the test, the population of respondents and the specifications for the test. The procedures used to develop the test must be clearly documented.

### Summary of the Guidelines

The test development process should be carefully and clearly documented. The tests developed for the study must be clear and accurate and include the directions, stimulus materials, and items. The tests must also be appropriate to the population of respondents and the specifications of the test. Each item of the test must meet appropriate technical standards. The tests must undergo extensive review processes, and a description should be made of how the items are to be scored and how scorers will be trained.

### Sources for this Standard

Adams, R. J. and Gonzalez, E. J. (1996) The TIMSS test design. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development* (pp. 3-1–3-36). Chestnut Hill, MA: TIMSS International Study Center.

In this chapter from the TIMSS 1995 technical report, Adams and Gonzales describe the booklet design for TIMSS tests for 1995. Their section on the constraints of the TIMSS test design is very detailed and could serve to guide future studies. The description of the cluster-based design, the organization of the test booklets, and the description of the test booklet contents are also very thorough.

Airasian, P. W. (1994) *Classroom Assessment*. New York: McGraw-Hill.

While this university textbook gives an excellent presentation of classroom assessment, Chapters 1, 5, 6 and 7 are especially pertinent to IEA-type studies. Airasian covers all aspects of testing in a clear manner, supporting key points with examples. He uses tables to provide excellent summaries of key points. For example, Table 6.02, 'Suggestions for Preparing Test Items', encapsulates the accumulated wisdom pertinent to writing good test items.

Airasian, P. W. (1996) *Assessment in the Classroom*. New York: McGraw-Hill.

This well-written textbook is aimed at the school practitioner. Chapter 3, 'Formal Assessment: Teacher-Made Tests', has a very good presentation of the considerations of test-making and test administration. The textbook as a whole serves as a useful introduction to the notion of school assessment.



Beaton, A. E. (1997) Item sampling. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.). Cambridge, UK: Pergamon.

Item sampling is concerned with the selection of items in the test or assessment instrument being used in measurement. The article considers the topics of matrix sampling, multiple matrix sampling, balanced incomplete block sampling, rotated sampling, the duplex design, and incidence sampling.

Cangelosi, J. S. (1990) *Designing Tests for Evaluating Student Achievement*. New York: Longman.

Chapter 4 of this university text, 'Item Development Hints', begins with five critical steps of item design. The author then proceeds to describe how to perform these steps when designing various types of items.

Garden, R. A. and Orpwood, G. (1996) Development of the TIMSS achievement tests. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development* (2-1-2-19). Chestnut Hill, MA: International Study Center.

In this chapter from the TIMSS 1995 technical report, Garden and Orpwood provide a detailed description of the process used to develop the TIMSS achievement tests. The authors include a timeline, details of field trials, linking items (between the three target populations), and guidelines on how to prepare for the main survey.

Garden, R. A. and Smith, T. A. (2000) TIMSS test development. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 49–70). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the development of the mathematics and science achievement tests for the TIMSS 1999 assessment.

Gronlund, N. E. (1998) *Assessment of Student Achievement* (6th ed.). Boston MA: Allyn and Bacon.

Gronlund's emphasis throughout this university textbook is on how to provide clear, practical guidelines for the assessment of student achievement. Chapter 7 describes how a test should be assembled, administered and evaluated. The first part of the chapter includes sections on reviewing and editing items, arranging items in the test, preparing directions and reproducing the test.

Herman, J. L. (1997) Item writing techniques. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 749–54). Cambridge, UK: Pergamon.

The article provides an overview of a range of current item-writing techniques. The author limits the overview to techniques for measuring academic achievement and focuses principally on selected response, or 'objective', measures.



Lie, S., Taylor, A. and Harmon, M. (1996) Scoring techniques and criteria. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development* (pp. 7-1–7-16). Chestnut Hill, MA: TIMSS International Study Center.

This chapter describes the development of diagnostic scoring rubrics for free-response items for the TIMSS 1995 assessments.

Millman, J. and Greene, J. (1989) The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 335–66). New York: American Council on Education and Macmillan Publishing.

This chapter, directed at the professional test constructor, emphasizes options for specifying and developing tests. The discussion begins with the purpose of the test, possible elements of a test plan, types of items and how they are written, and the evaluation and selection of items. It concludes with a brief section on putting the items together into a test.

Ory, J. C. and Ryan, K. E. (1993) *Tips for Improving Testing and Grading* (Survival Skills for Scholars). Newbury Park, CA: Sage Publications.

This book is aimed at the school-practitioner. Chapters 3 and 4 have suggestions for writing test items. Both chapters include examples, and the major points are clearly presented.

Tittle, C. K. (1997) Test bias. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 813–19). Cambridge, UK: Pergamon.

Test bias is defined as being prejudicial or unfair to groups or individuals who differ from the majority of test-takers. This article presents an overview of test bias, including test bias methodology.





## Standard for Questionnaire Development

**Purpose** To ensure that questionnaires developed for IEA studies address the issues specified in the study questions in a way that maximizes the reliability and validity of the measures while minimizing the burden on respondents.

**Standard** Questionnaires should be clear, simple, concise and manageable.

**Summary of the Guidelines** Questionnaires should be clear, simple, concise and manageable. When designing questionnaires, the goal is to clearly define the information needed, and to develop as few questions (preferably of a closed format) as possible to obtain the information.

**Sources for this Standard** Aiken, L. R. (1997) *Questionnaires and Inventories: Surveying Opinions and Assessing Personality*. New York: John Wiley and Sons.

*Questionnaires and Inventories* is written for researchers and practitioners in the above-mentioned disciplines and for all others who rely upon these important tools-of-the-trade. After concisely reviewing the historical origins and theoretical underpinnings of assessment tools and rating scales, Aiken presents chapters providing detailed, hands-on coverage of the construction, administration, scoring and interpretation of questionnaires, psychological inventories, and population surveys. Subsequent chapters feature in-depth discussions of the statistical analysis of data, as well as test reliability and validity. The final two chapters are devoted to the critical assessment of most commercially available questionnaires and inventories.

Converse, J. M. and Presser, S. (1995) *Survey Questions: Handcrafting the Standardized Questionnaire* (Quantitative Applications in the Social Sciences, Vol. 63). Thousand Oaks, CA: Sage Publications.

*Survey Questions* is a guide to the principles of writing survey questions. The authors review recent research on survey questions, consider the lore of professional experience, and present those findings that have the strongest implications on writing survey questions.

Fowler, F.J. (1995) *Improving Survey Questions: Design and Evaluation* (Applied Social Research Methods Series, Vol. 38). Thousand Oaks, CA: Sage Publications.

This book shows how to word and format questions that will evoke the kinds of answers for which they are designed. In addition, the book covers the following: how to write good questions aimed at collecting information about objective facts and events; measuring subjective phenomena; some alternative methods for attacking common measurement problems; how to evaluate the extent to which questions are consistently understood and administered; and how to evaluate the data resulting from a set of questions.



Keeves, J. P. and Saha, L. J. (1997) Measurement of social background. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 930–44). Cambridge, UK: Pergamon.

This article is concerned with the measurement of social background and social status, and the problems encountered in the development of consistent and valid measurements.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

This practical book is designed for people new to research and planning who intend to undertake a research project. Chapter 10, 'Collecting Data Using Attitudinal Scales', describes how scales can be developed to assess attitudes.

Mullis, I. V. S., Martin, M. O. and Stemler, S. E. (2000) TIMSS questionnaire development. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 71–88). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the development of the TIMSS questionnaires used in 1999.

Oppenheim, A. N. (1992) *Questionnaire Design, Interviewing, and Attitude Measurement*. New York: Pinter Publishers.

*Questionnaire Design* provides a comprehensive coverage of survey design and implementation. The chapter on measurement is limited to classical measurement theory.

Schmidt, W. H. and Cogan, L. S. (1996) Development of the TIMSS context questionnaires. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development*. Chestnut Hill, MA: TIMSS International Study Center.

This chapter from the TIMSS 1995 technical report describes the conceptual models and questionnaire development process for TIMSS 1995.

Schuman, H. and Presser, S. (1981) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. San Diego, CA: Academic Press.

This comprehensive book, which covers aspects of attitude surveys, includes question order, tone of questions, open versus closed questions, and problems with respondent acquiescence. The authors introduce concepts and promote in-depth understanding through their use of concrete examples.

Sudman, S. and Bradman, N. M. (1982) *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco, CA: Jossey-Bass.

This very thorough and easy-to-read book leads a person new to the field through the concepts of questionnaire design. Each chapter has a checklist of major points, and all aspects of questionnaire design are covered. The authors use examples to reinforce the main concepts.



Sudman, S., Norman M., Bradburn, N. M. and Schwarz, N. (1996) *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass.

This textbook is primarily aimed at helping survey research practitioners write better questionnaires. Chapters cover a wide range of concepts, including methods for determining cognitive processes involved when respondents answer a question, the difficulties caused by the research situation, the psychological sources of context effects, and order effects within a question.

Tanur, J. M. (Ed.) (1992) *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.

This book contains chapters covering a wide range of topics pertinent to questionnaire design. Perhaps the most relevant to IEA surveys are those contained in Chapter 2, 'Asking Questions and Influencing Answers', and in all the chapters of Part 4, 'Expression: The Case of Attitude Measurement in Surveys'.

Wolf, R. M. (1997) Questionnaires. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 422–27). Cambridge, UK: Pergamon.

Wolf, in this very good overview of questionnaire development, describes the assumptions associated with use of questionnaires, the steps in developing a questionnaire, the pilot stage, how to refine the questionnaire, field testing, validity and reliability, and coding systems.



## Standard for Translations and Verifying Translations

### Purpose

To ensure that (i) the cognitive items are translated from the international versions into the target languages without changes in meaning or difficulty; (ii) that cultural differences are kept to a minimum; and (iii) that the meaning and content of the questionnaire items are retained through translation. The goal is to obtain translated instruments of high quality that will provide comparable data across countries and cultures.

### Standard

When translating test items or modifying them for cultural adaptation, the following must remain the same as the international version: the meaning of the question; the reading level of the text; the difficulty of the item; and the likelihood of another possible correct answer for the test item.

### Summary of the Guidelines

When translating test items or modifying them for cultural adaptation, the following must remain the same as the international version: the meaning of the question; the reading level of the text; the difficulty of the item; and the likelihood of another possible correct answer for the test item.

### Sources for this Standard

Danks, J. H., Shreve, G. M., Fountain, S. B. and McBeath, M. (Eds.) (1997) *Cognitive Processes in Translation and Interpreting Applied Psychology: Individual, Social and Community Issues* (Vol. 3). Thousand Oaks, CA: Sage Publications.

The interdisciplinary contributions in this volume provide a strong foundation for improvement in translation skills. The authors apply concepts and methods of cognitive science to translation, focusing on the relationship between translation theory, research and practice.

Litwin, M. S. (1995) *How to Measure Survey Reliability and Validity Series* (The Survey Kit, Vol. 7). Thousand Oaks, CA: Sage Publications.

Litwin explains how to code and pilot-test new and established surveys. In addition, he covers such issues as the following: how to measure reliability (including test-retest, alternate form, internal consistency, inter-observer, and intra-observer reliability); how to measure validity (including content, criterion, and construct validity); how to address cross-cultural issues in survey research; and how to scale and score a survey. In Chapter 7, 'Multicultural Issues', Litwin raises the need to be sensitive to multicultural issues in survey design.



Maxwell, B. (1996) Translation and cultural adaptation of the survey instruments. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development* (pp. 8-1–8-8). Chestnut Hill, MA: TIMSS International Study Center.

This chapter from the TIMSS 1995 technical report describes the processes used to adapt the international version of the 1995 TIMSS questionnaires to suit different language and cultural settings.

McKay, R. B., Breslow, M. J., Sangster, R. L., Gabbard, S. M., Reynolds, R. W., Nakamoto, J. M. and Tarnai, J. (1996) Translating survey questionnaires: lessons learned. In M. T. Braverman and J. K. Slater (Eds.), *Advances in Survey Research* (pp. 93–104). San Francisco, SF: Jossey-Bass.

The authors discuss factors that help make a translated instrument comparable to the original survey instrument as well as more comprehensive to the respondent.

Mullis, I. V. S., Kelly, D. L. and Haley, K. (1996) Translation verification. In M. O. Martin and I. V. S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection* (pp. 1-1–1-13). Chestnut Hill, MA: TIMSS International Study Center.

This chapter from the TIMSS 1995 technical report documents the procedures TIMSS used to verify the accuracy of the translations of the TIMSS instruments. Specifically, the TIMSS instrument translation and verification included explicit guidelines for translation and cultural adaptation, verification of the quality of the translations and booklet layout by independent translators, and statistical checks after testing to detect items that did not perform comparably across countries.

O'Connor, K. M. and Malak, B. (2000) Translation and cultural adaptation of the TIMSS instruments. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 89–102). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the translation and translation verification of the TIMSS 1999 data collection instruments.



## Standard for Field Testing Data Collection Instruments and Procedures

### Purpose

To obtain information about the sampling procedures, the test administration procedures, the performance of the test questions and the scoring procedures before the questions are included in the final forms and the procedures are used in the actual assessment. Note: the purpose is to field test<sup>1</sup> the items and procedures, not the students.

### Standard

Here, the aims are to:

- (i) Field test all items, tasks and directions on a population that is as similar as possible to the target population.
- (ii) Use as many as possible of the sampling, test administration and scoring procedures that are planned for the actual data collection phase of the study.
- (iii) Use smaller scale field-testing approaches as appropriate before field testing to ensure that the full-scale field test provides the most accurate information possible.

### Summary of the Guidelines

The assessment instruments should be administered to a sample from a population in each participating country that is similar to the national target population. The field test should involve sufficient items to permit assembly of the final forms from the field-tested items. The items should then be submitted to technical analyses to detect flawed or biased items. The performance of questionnaire items should also be examined.

### Sources for this Standard

Camilli, G. and Shepard, L. A. (1994) *Methods for Identifying Biased Test Items* (Measurement and Methods in the Social Sciences, Vol. 4). Thousand Oaks, CA: Sage Publications.

This book makes clear to researchers what item-bias methods can (and cannot) do, how they work, and how they should be interpreted. Advice is provided on the most useful methods for particular test situations. The authors explain the logic of each method—from item-response theory to non-parametric, categorical methods—in terms of how differential item functioning (DIF) is defined by the method and how well the method can be expected to work. They also provide a summary of findings on the behavior of indices in empirical studies, and conclude with a set of principles for deciding when DIF should be interpreted as evidence of bias. The book is ideal for those with a basic understanding of measurement and statistics, and those who wish to gain an understanding of item bias.

<sup>1</sup> In this standard, any type of experimental tryout or trial of new materials before the administration of the final form is referred to as field testing. Field testing can take many forms from 'think alouds' with 10 or so students to help gain a better understanding of new item types, through pilot testing with a small group of countries, to full-scale field testing that simulates the procedures planned for the final assessment as closely as possible.



Garden, R. A. and Orpwood, G. (1996) Development of the TIMSS achievement tests. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development* (pp. 2-1–2-19). Chestnut Hill, MA: TIMSS International Study Center.

In this chapter from the TIMSS 1995 technical report, Garden and Orpwood include a detailed description of the process used to field test the TIMSS achievement tests. The authors include a timeline, details of field trials, linking items (between the three target populations), and guidelines on how to prepare for the main survey.

O'Connor, K. M. (2000) TIMSS field test. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 103–18). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes in detail the field test for the TIMSS 1999 assessment, including the field test booklet design, the data collection procedures and the analysis of the field test data.

Ory, J. C. and Ryan, K. E. (1993) *Tips for Improving Testing and Grading* (Survival Skills for Scholars). Newbury Park, CA: Sage Publications.

This book is aimed at the school practitioner. Chapter 6, 'Evaluating the Quality of Classroom Exams', briefly addresses the methods used to evaluate examinations, including item and test review, and the use of item statistics.



# Standards for Data Collection and Processing

## Standard for Drawing a Sample

### Purpose

To ensure that the samples drawn in each country comply with the study sampling plan and are representative of the specified target population, with minimal risk of bias, and that the sampling procedure will accommodate the computation of sampling weights and variances.

### Standard

Samples for IEA studies must conform to the study sampling design and use methods derived from sound and defensible sampling theory. They must meet specified quality standards in terms of coverage, participation rates and data reliability.

### Summary of the Guidelines

Samples for IEA studies must conform to the study sampling design and use methods derived from sound and defensible sampling theory. They must meet quality standards in terms of coverage, participation rates and data reliability.

### Sources for this Standard

Fink, A. (1995) *How to Sample in Surveys* (The Survey Kit, Vol. 6). Thousand Oaks, CA: Sage Publications.

How to select and use the most appropriate sampling methods for a survey is the focus of this book. It covers a range of sampling techniques that include simple random sampling, stratified random sampling, systematic sampling, cluster sampling, convenience sampling, snowball sampling, quota sampling, and focus groups. The book also describes in detail the following: inclusion and exclusion criteria, the logic behind estimating standard errors and sample size determinations, the sources of error in sampling, and how to calculate response rates.

Foy, P. and Joncas, M. (2000) Implementation of the sample design. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 157–70). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the sampling design for TIMSS 1999.





Foy, P., Rust, K. and Scheicher, A. (1996) Sample design. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 1: *Design and Development* (pp. 4-1-4-17). Chestnut Hill, MA: TIMSS International Study Center.

Foy, Rust and Scheicher describe the TIMSS 1995 sample design in this chapter from the technical report. They describe the procedures developed to ensure that the student populations were properly sampled within each country.

Lee, E. S., Forthofer, R. N. and Lorimer, R. J. (1989) *Analyzing Complex Survey Data*. Newbury Park, CA: Sage Publications.

*Analyzing Complex Survey Data* offers an effective means of examining and analyzing complex surveys, and provides guidance on how to overcome the problems that often arise. It includes discussions on the choices involved in variance estimates, simple random sampling without replacement, stratified random sampling, and two-stage cluster sampling. It also outlines the different computer programs that are currently available.

Ross, K. N. and Rust, K. (1997) Sampling in survey research. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 427-38). Cambridge, UK: Pergamon.

This article on sample design for educational research focuses upon some aspects of the survey approach to research and its application to large-scale assessment.



## Standard for Planning for Data Collection

### Purpose

To ensure that comparable data are collected in participating countries while recognizing national variations in school system characteristics and data availability.

### Standard

Planning for data collection involves implementing the international study design in the form of standardized procedures, forms and materials to ensure comparability of the data collected. All aspects of data collection should be specified in detail in the plan, including: selecting respondents and securing co-operation; hiring and training study staff; documenting survey procedures, questionnaires and forms; and specifying quality assurance procedures.

### Summary of the Guidelines

The development of a detailed data collection plan should involve the representatives of the participating countries. The plan should be used to develop manuals that specify the procedures and materials appropriate for each level of data collection.

### Sources for this Standard

Gonzalez, E. J. and Hastedt, D. (2000) TIMSS field operations and data preparation. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 119–34). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the procedures for data collection for the TIMSS 1999 assessment.

Schleicher, A. and Siniscalco, M. T. (1996) Field operations. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report. Volume 1: Design and Development* (pp. 9-1–9-25). Chestnut Hill, MA: TIMSS International Study Center.

The field operations described in this chapter from the TIMSS 1995 technical report are based on procedures used successfully in previous IEA studies and enhanced on the basis of the TIMSS field trial. The chapter presents the design of the field operations from the first contact with the sampled schools to the return of the cleaned data files to the IEA data processing center.



## Standard for Selecting and Training Data Collection Staff

**Purpose** To make sure that data collection staff have the knowledge and skills to carry out all data collection activities.

**Standard**

The qualifications required of data collection staff should reflect the complexity of the tasks they will perform. Training programs in data collection procedures should be sufficiently thorough that trainees complete their training with all the skills necessary to carry out their responsibilities successfully. Training staff should be able to evaluate the progress of each trainee as they work through the training program.

**Summary of the Guidelines**

The qualifications required of data collection staff should reflect the complexity of the tasks they will perform. Training programs in data collection procedures should be sufficiently thorough to allow trainees to complete their training with all the skills necessary to carry out their responsibilities successfully. Training staff should be able to evaluate the progress of each trainee as they work through the training program.

**Sources for this Standard**

Forsyth, D. R. (1990) *Group Dynamics* (2nd ed.). Pacific Grove, CA: Brooks/Cole Publishing.

*Group Dynamics* is a graduate-level psychology/sociology textbook. This textbook has relevance to this standard because study personnel benefit from training that facilitates an understanding of group behavior and group leadership. The chapters on group leadership (Chapter 3), group structure (Chapter 5), leadership (Chapter 8) and performance (Chapter 9) are all relevant to study center operations. For example, Chapter 9 has sections on predicting potential group performance, productivity losses in groups and productivity gains in groups.

Härnqvist, K. (1997) Training of research workers in education. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.) (pp. 269–76). Cambridge, UK: Pergamon.

This article describes the special demands made of training staff involved in large-scale assessment. It also looks at the various problems involved in the design and implementation of appropriate training programs.



## Standard for Minimizing Response Burden and Non-response

### Purpose

To ensure that data collection activities are minimally intrusive and burdensome and that the data collected represent the population being studied.

### Standard

Since burden and non-response are related to each other, and data collection procedures and materials are based on the study design, all aspects of the study design should be carefully reviewed against realistic expectations of respondents, recognizing that participation is generally voluntary and not a primary activity for respondents.

As materials and procedures are developed to implement the design, they should be reviewed for clarity, ease of use and simplicity. Acceptable levels of response and the role of replacements should be defined before beginning the data collection. Standards for reporting data should also be defined before the beginning of data collection.

### Summary of the Guidelines

Data collection activities should be minimally intrusive and burdensome, and the data collected should represent the population being studied.

### Sources for this Standard

Krosnick, J. A., Narayan, S. and Smith, W. R. (1996) Satisficing in surveys: initial evidence. In M. T. Braverman and J. K. Slater (Eds.), *Advances in Survey Research* (pp. 29–24). San Francisco, CA: Jossey-Bass.

This article examines the cognitive burden a respondent experiences, the consequences of such a burden, and the methods that can be used to prevent or minimize it.

Lessler, J. T. and Kalsbeek, W. D. (1992) *Nonsampling Error in Surveys*. New York: John Wiley and Sons.

The authors focus on three areas of non-sampling survey errors: frame, non-response and measurement. Each one is analyzed accordingly: definition of key terms, formulation of known effects, and examination of suggested remedies. The terminology related to each source of error is reviewed.



## Standard for Implementing Data Collection Quality Control Procedures

<b>Purpose</b>	To make sure that data are collected according to study requirements.
<b>Standard</b>	<p>Quality control should be an integral part of the study at both the national and international levels. Quality control encompasses internal mechanisms that are built into each stage of data collection to ensure that procedures are implemented correctly, and external reviews administered by staff members who are separate from the staff being evaluated.</p>
<b>Summary of the Guidelines</b>	Quality control should be an integral part of the study both at the national and international levels.
<b>Sources for this Standard</b>	<p>Martin, M. O., Hoyle, C. D. and Gregory, K. D. (1996) Monitoring the TIMSS data collection. In M.O. Martin and I. V. S. Mullis (Eds.), <i>Third International Mathematics and Science Study: Quality Assurance in Data Collection</i> (pp. 3-1–3-13). Chestnut Hill, MA: TIMSS International Study Center.</p> <p>This chapter describes the TIMSS 1995 quality assurance program. It refers to the training of quality control monitors, the role of these monitors, and how the information gathered by the monitors was used.</p> <p>Martin, M. O., Hoyle, C. D. and Gregory, K. D. (1996) Observing the TIMSS test administration. In M. O. Martin and I. V. S. Mullis (Eds.), <i>Third International Mathematics and Science Study: Quality Assurance in Data Collection</i> (pp. 4-1–4-10). Chestnut Hill, MA: TIMSS International Study Center.</p> <p>As part of the 1995 data collection, TIMSS sent quality control monitors to a sample of schools where they observed testing sessions and interviewed school coordinators. This chapter from the quality assurance report describes the school visits and test session observations.</p> <p>Naus, J. I. (1975) <i>Data Quality Control and Editing, Statistics: Textbooks and Monographs</i> (Vol. 10). New York: Marcel Dekker.</p> <p>The section on quality control procedures describes procedures for detecting errors in sets of data. The author discusses the advantages and disadvantages as well as the methods of automatic correction.</p>



O'Connor, K. M. and Stemler, S. E. (2000) Quality control in the TIMSS data collection. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 135–56). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the program of school visits during the TIMSS 1999 data collection, and summarizes the observations made by the quality control monitors.



## Standard for Documenting National Data Collection

<b>Purpose</b>	To ensure that the national data collection activities can be reviewed by the international coordinating center (ICC) and documented for use by later users of the data.
<b>Standard</b>	The documentation of the data collection for each country should include all adaptations to and deviations from the international procedures that have been made by the country.
<b>Summary of the Guidelines</b>	The documentation of the data collection for each country should include all adaptations to and deviations from the international procedures that have been made by the country.
<b>Source for this Standard</b>	<p>Kelly, D. L., Mullis, I. V. S. and Smith, T. A. (1997) Reporting student and teacher questionnaire data. In M. O. Martin and D. L. Kelly (Eds.), <i>Third International Mathematics and Science Study Technical Report. Volume 2: Implementation and Analysis</i> (pp. 198–203). Chestnut Hill, MA: TIMSS International Study Center.</p> <p>This chapter from the TIMSS 1995 technical report describes the national adaptations made by participating countries to the TIMSS 1995 questionnaires, and the steps taken to deal with those adaptations during the reporting of results.</p>



## Standard for Planning for Data Preparation and Processing

### Purpose

To ensure that all participants have the tools and information that they need for data preparation and processing so that (i) the participants are ready to produce data files for the international coordinating center (ICC) that conform to the international data structure, and (ii) the tools and information are comparable internationally and as free from error as possible.

### Standard

National coordinators should plan their data preparation activities in good time, and ensure that they have all of the information and understand the procedures necessary to create data files that conform to the international data structures. The ICC should ensure that participants understand the importance of planning for effective data preparation, and should provide them with all the necessary tools, information and training in data preparation. All prescribed data preparation procedures and quality control checks should be clearly documented. Training in the use of software and in the application of data preparation procedures should be provided, as necessary.

### Summary of the Guidelines

The national coordinators should ensure that they fully understand all data preparation and processing procedures. The ICC should provide participants with manuals describing all aspects of data entry, including the assignment of identification numbers, data formats and file structures. The ICC should provide training and assistance to national coordinators.

### Sources for this Standard

Bourque, L. B. and Clark, V. A. (1992) *Processing Data: The Survey Example* (Quantitative Applications in the Social Sciences, Vol. 85). Thousand Oaks, CA: Sage Publications.

This volume highlights the theory that decisions made during the design of a data collection instrument influence the kind of data and the format of the data that are available for analysis. Opening with a discussion on the selection of the data collection technique(s) and how this impacts on data processing and the data for later analysis, the book covers key issues, such as the following: whether a researcher should create his or her own instrument for a questionnaire; how to test a questionnaire; the characteristics of good data processing; how to deal with missing data; and how to scale an evaluation and create sub-files for analysis. In addition, each major section concludes with examples and, when appropriate, directs the reader to commonly available computer software that can aid in data processing.





Coxon, A. P. M. (1998) *Sorting Data: Collection and Analysis* (Quantitative Applications in the Social Sciences, Vol. 127). Thousand Oaks, CA: Sage Publications.

Coxon gives hints on how to collect, describe, compare and analyze data. The author addresses specification, elaboration, and sampling of the 'domain' or what is to be sorted. There is also help on setting the criterion, the pre-test, administration, and recording of results. The author gives special consideration to problems of categorization, illustrating this with a real research example.

Fowler, F. J. (1993) *Survey Research Methods* (Applied Social Research Methods Series, Vol. 1, 2nd ed.). Thousand Oaks, CA: Sage Publications.

This edition of *Survey Research Methods* covers the role of microcomputers in data collection and data entry procedures.



## Standard for Preparing Data for Processing and Checking Data

<b>Purpose</b>	To ensure that data provided by national centers are in a common format and are suitable for international analyses.
<b>Standard</b>	<p>The data provided by national centers should conform to the international data structure and to the data definitions provided in the international code book. They should also be free of internal inconsistencies. The international coordinating center (ICC) should ensure that all data files provided by participating countries comply with international formats and are free from error.</p>
<b>Summary of the Guidelines</b>	<p>National coordinators should ensure that not only all prescribed procedures for data preparation but also all prescribed checks for data integrity and quality control are closely followed. The ICC should encourage and support the national coordinators in their endeavors, check the data sets provided by the participants against the international data structure and report all deviations to the providers. The ICC should also report all extreme or unlikely values to the national centers, and provide review item analyses and preliminary univariate statistics to each national center.</p>
<b>Sources for this Standard</b>	<p>Davidson, F. (1996) <i>Principles of Statistical Data Handling</i>. Thousand Oaks, CA: Sage Publications.</p> <p><i>Principles of Statistical Data Handling</i> will help the reader understand the principles of data handling and make better use of computer data in research or study. The book demonstrates how to input, manipulate and debug data to make substantive analysis easier and more accurate. Using a series of universal concepts, Davidson presents a situation or a problem, suggests how it might be resolved, and demonstrates the implementation of each principle as it appears in the command languages of SAS and SPSS.</p> <p>Mullis, I. V. S., Jones, C. and Garden, R. A. (1996) Training sessions for free-response scoring and administration of performance assessment. In M. O. Martin and D. L. Kelly (Eds.), <i>Third International Mathematics and Science Study Technical Report</i>. Volume 1: <i>Design and Development</i> (pp. 10-1–10-12). Chestnut Hill, MA: TIMSS International Study Center.</p> <p>This chapter from the TIMSS 1995 technical report describes the training sessions used to enable people to score the TIMSS free-response items and to administer the performance assessment.</p>



Mullis, I. V. S. and Smith, T. A. (1996) Quality control steps for free-response scoring. In M. O. Martin and I. V. S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection* (pp. 5-1-5-32). Chestnut Hill, MA: TIMSS International Study Center.

TIMSS in 1995 used a three-pronged approach to meet the goal of reliable scoring. First, TIMSS conducted training sessions for representatives from each country on the use of the two-digit free-response coding. Second, TIMSS gathered and documented information about the within-country agreement among coders. Third, TIMSS conducted a special study where coders from the TIMSS countries coded common sets of student responses.

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

After establishing that only standard complete-data methods are needed to analyze a multiply imputed set, the author evaluates procedures in general circumstances, outlining specific procedures for creating imputations in both the ignorable and non-ignorable cases. Examples and exercises reinforce ideas, and the interplay of Bayesian and frequentist ideas presents a unified picture of modern statistics.



## Standard for Documenting Data-Processing Activities

### Purpose

To ensure that all modifications applied by the international coordinating center (ICC) to participants' data because of structural reorganization or error detection are recorded, so that the integrity of the data may be clearly demonstrated and all deviations from the international standard explicitly documented.

### Standard

National coordinators should be notified of all errors and inconsistencies encountered in their data. All modifications to data arising from the resolution of such issues should be endorsed by the national research center and clearly documented, preferably in a database. National deviations from international standards should also be clearly documented, to ensure the validity of international comparisons and to facilitate subsequent secondary analysis by researchers who do not have first-hand experience of the data collection.

### Summary of the Guidelines

National coordinators should be notified of all errors and inconsistencies encountered in their data. All modifications should be endorsed by the national research center and clearly documented, preferably in a database. National deviations from international standards should also be clearly documented to ensure the validity of international comparisons and to facilitate subsequent secondary analysis by researchers who do not have first-hand experience of the data collection.

### Sources for this Standard

Hastedt, D. and Gonzalez, E. J. (2000) Data management and database construction. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 171–88). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the data checking and database construction for TIMSS 1999.

Jungclaus, H. and Bruneforth, M. (1996) Data consistency checking across countries. In M. O. Martin and I. V. S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection* (pp. 6-1–6-9). Chestnut Hill, MA: TIMSS International Study Center.

This chapter describes the quality-control procedures used for checking the 1995 TIMSS data. More specifically, it describes the steps that were involved in cleaning the data, and in standardizing the structure of the files across countries, and the procedures implemented to facilitate the construction of the international database.



Mullis, I. V. S. and Smith, T. A. (1996) Quality control steps for free-response scoring. In M.O. Martin and I. V. S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection* (pp. 5-1-5-32). Chestnut Hill, MA: TIMSS International Study Center.

TIMSS used a three-pronged approach to meet the goal of reliable scoring. First, TIMSS conducted training sessions for representatives from each country on the use of the two-digit free-response coding. Second, TIMSS gathered and documented information about the within-country agreement among coders. Third, TIMSS conducted a special study where coders from the TIMSS countries coded common sets of student responses.

Sibberns, H., Hastedt, D., Bruneforth, M., Schwippert, K. and Gonazalez, E. J. (1997) Data management and construction of the TIMSS database. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report. Volume 2: Implementation and Analysis* (pp. 47-67). Chestnut Hill, MA: TIMSS International Study Center.

This chapter from the TIMSS 1995 technical report provides an overview of the data-processing activities for TIMSS 1995 at the IEA data processing center in Hamburg.



# Analyzing Data and Reporting Results

## Standard for Developing Sampling Weights

<b>Purpose</b>	To assign sampling weights to all sampled units so that appropriate population inferences can be made.
<b>Standard</b>	Sampling weights that accurately reflect the probability of selection should be assigned to each sampled unit.
<b>Summary of the Guidelines</b>	Sampling weights that accurately reflect the probability of selection should be assigned to each sampled unit.
<b>Sources for this Standard</b>	<p>Foy, P. (1997) Calculation of sampling weights. In M. O. Martin and D. L. Kelly (Eds.), <i>Third International Mathematics and Science Study Technical Report. Volume 2: Implementation and Analysis</i> (pp. 71–80). Chestnut Hill, MA: TIMSS International Study Center.</p> <p>This chapter from the TIMSS 1995 technical report describes the calculation of sampling weights for the TIMSS 1995 data.</p> <p>Foy, P. (2000) Sampling weights. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), <i>Third International Mathematics and Science Study: 1999 Technical Report</i> (pp. 189–202). Chestnut Hill, MA: Boston College.</p> <p>This chapter from the TIMSS 1999 technical report describes the calculation of sampling weights for the TIMSS 1999 data.</p> <p>Kish, L. (1965) <i>Survey Sampling</i>. New York: John Wiley and Sons.</p> <p>Kish's excellent textbook on survey sampling provides a very good methodological background to the TIMSS' applications provided by Foy.</p>



## Standard for Reporting Sampling and Non-sampling Errors

### Purpose

To ensure that data analysts have all the information necessary to make informed decisions about the quality and reliability of the data that they are using.

### Standard

Estimates of sampling error should be presented for all descriptive statistics included in IEA reports. Data quality measures on potential sources of non-sampling error, such as response rates, population coverage and exclusions, also should be reported.

### Summary of the Guidelines

Estimates of sampling error should be presented for all descriptive statistics included in IEA reports. Data quality measures on potential sources of non-sampling error, such as response rates, population coverage and exclusions, also should be reported.

### Sources for this Standard

Gonzalez, E. J. and Foy, P. (1997) Estimation of sampling variability, design effects and effective sample sizes. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report*. Volume 2: *Implementation and Analysis* (pp. 81–86). Chestnut Hill, MA: TIMSS International Study Center.

This chapter from the TIMSS 1995 technical report describes the jackknife repeated replication (JRR) and balanced repeated replication (BRR) methods used to estimate sampling variability in the TIMSS 1995 data.

Gonzalez, E. J. and Foy, P. (2000) Estimation of sampling variance. In M. O. Martin, K. D. Gregory and S. E. Stemler (Eds.), *Third International Mathematics and Science Study: 1999 Technical Report* (pp. 203–24). Chestnut Hill, MA: Boston College.

This chapter from the TIMSS 1999 technical report describes the jackknife repeated replication (JRR) procedures used to estimate sampling variability in TIMSS 1999.

Kish, L. (1965) *Survey Sampling*. New York: John Wiley and Sons.

Kish's classic book on sampling techniques devotes considerable attention to analytic techniques for estimating sampling errors in both simple and complex sample designs.



Longford, N. T. (1995) *Models of Uncertainty in Educational Tests*. New York: Springer-Verlag.

This excellent technical book is suitable for people at ease with statistics. The book presents a number of statistical techniques applicable to advanced survey design. In Chapter 7, 'Inferences from Surveys with Complex Sampling Design', Longford explores issues in sampling design (the realized sample design, the model sampling design, sampling weights, and non-response). Other techniques that the author considers include imputed values, jackknifing, stratification and clustering, and weights. The chapter represents a very comprehensive and easy-to-read technical presentation that is pertinent to IEA-type studies.

Ross, K. N. and Wilson, M. (1997) Sampling errors in survey research. In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (pp. 663–70). Cambridge, UK: Pergamon.

This article considers the issues involved in the estimation of sampling errors and variance, the empirical procedures employed in their estimation, and developments to model the structure of the sample design and to improve the estimation of variance components.





## Standard for Validating Constructs and Scales for Analysis

### Purpose

To ensure that all measurement scales (both cognitive and non-cognitive) used in analyses are valid measures of the intended constructs so that consumers of study reports may have confidence in the validity of the results.

### Standard

Every effort should be made to establish the validity and reliability of all measurement scales reported in IEA studies. The content validity of measures of achievement should always be demonstrated. Similarly, some evidence of construct validity is also necessary for scales purporting to measure attitudes, opinions or other constructs. Where appropriate, evidence of the reliability of all scales should be provided in international reports.

### Summary of the Guidelines

All measurement scales used in analyses should be valid measures of the intended constructs so that consumers of study reports may have confidence in the validity of the results.

### Sources for this Standard

DeVellis, R. F. (1991) *Scale Development: Theory and Applications* (Applied Social Research Methods Series, Vol. 26). Newbury Park, CA: Sage Publications.

This short book provides an easy-to-read account of scale development, reliability and validity, and clearly illustrates eight steps for scale development.

Eisenhart, M. A. and Howe, K. R. (1992) Validity in educational research. In M. S. LeCompte, W. L. Millroy and J. Preissle (Eds.), *The Handbook of Qualitative Research in Education*. San Diego, CA: Academic Press.

The authors provide a scholarly treatise on validity in educational research from a qualitative research perspective. They cover the conventional conception of validity as defined by Campbell and Stanley, and discuss several kinds of alternative conceptions that grew out of, and responded to, the special features of qualitative research.

Holland, P. W. and Wainer, H. (Eds.) (1993) *Differential Item Functioning*. Mahwah, NJ: Lawrence Erlbaum.

Test fairness is a moral imperative for both the makers and users of tests. This book focuses on methods for detecting test items that function differently for different groups of examinees and on ways to use this information to improve tests. Of interest to all testing and measurement specialists, it examines modern techniques used to ensure test fairness. Three of these relevant to the book's contents are: detailed reviews of test items by subject-matter experts and members of the major subgroups in society (gender, ethnic and linguistic) that will be represented in the examinee population; comparisons of the predictive validity of the test done separately for each of the major subgroups of examinees; and extensive statistical analyses of the relative performance of major subgroups of examinees on individual test items.



McMillan, J. H. (2000) *Educational Research: Fundamentals for the Consumer* (3rd ed.). New York: Longman.

This book is aimed at university students and seeks to make them intelligent consumers of education research. Chapters 6 and 7, on educational measurement, include the latest research on reliability and validity.

Traub, R. E. (1994) *Reliability for the Social Sciences: Theory and Applications* (Measurement and Methods in the Social Sciences, Vol. 3). Thousand Oaks, CA: Sage Publications.

Through an illustrated review of the principles of classical reliability theory, Traub explores some general strategies for improving measurement procedures. Beginning with a presentation of random variables and the expected value of a random variable, he covers such topics as the definition of reliability as a coefficient and possible uses of a coefficient; the notion of parallel tests so as to make possible the estimation of a reliability coefficient for a set of measurements; what to do when parallel tests are not available; what factors affect the reliability coefficient; and how to estimate the standard error of measurement. Aimed at giving readers a non-technical treatment of classical reliability theory, the book also includes end-of-chapter exercises as well as boxes that give more in-depth coverage of major topics or that provide algebraic proofs.



## Standard for Presenting Findings

### Standard

To facilitate widespread use of the study results and valid interpretation of the findings as they relate to the intended purposes of the study (see also **Standard on Reviewing the Primary Reports of Study Findings**).

### Purpose

Here, the requirement is to ensure that the IEA study findings are presented clearly, accurately and in a timely fashion. To increase their use, study findings should be made available to different audiences in the forms most useful for those audiences. Note that even though this standard applies most directly to the presentation of study findings in publications sponsored directly by the IEA, plans for dissemination can also include other vehicles, such as presentations at professional meetings, papers for scientific journals and articles for professional magazines.

### Summary of the Guidelines

The IEA study findings should be presented clearly, accurately and in a timely fashion, and in a form most useful to each audience.

### Sources for this Standard

Anastas, J. W. (1994) *Research Design for Social Work and the Human Services*. New York: Lexington Books.

While not specifically centered upon the educational setting, most of the concepts covered are pertinent to educational research. The section on research ethics is particularly good, as are the chapters on sampling, basic issues in data collection, and questionnaire design. The final chapter, 'Writing Research Reports', has good, clear guidelines for report writing.

Barnet, S., Stubbs, M. and Bellanca, P. (2000) *The Practical Guide to Writing with Readings and Handbook* (8th ed.). New York: Longman.

Stressing writing as a process involving both reading and critical thinking, this textbook helps students create skillful essays in exposition, analysis and persuasion.

Charles, C. M. (1998) *Introduction to Educational Research* (3rd ed.). New York: Longman.

This is an introductory textbook for students interested in understanding, conducting and interpreting both qualitative and quantitative paradigms in educational research methods. It focuses on eight types of quantitative and qualitative research, and includes selections of non-technical, reprinted research articles to orient students who wish to conduct similar studies. Part 3, 'Conducting Your Own Research Project', includes a chapter on preparing a research report.



Chow, S. L. (1996) *Statistical Significance: Rationale, Validity, and Utility*. Thousand Oaks, CA: Sage Publications.

Beginning with a thorough introduction to null hypothesis testing and statistical significance, the book advances the arguments for and against current interpretations and uses of significance testing in research.

Clines, R. H. and Cobb, E. R. (2000) *Research Writing Simplified* (3rd ed.). New York: Longman.

Designed to augment any university course requiring writing and the use of secondary source material, *Research Writing Simplified* offers students a set of activities and exercises that provide sequenced practice in the conventions of documentation.

Davis, M. (1996) *Scientific Papers and Presentations*. San Diego, CA: Academic Press.

*Scientific Papers and Presentations* gives a concise guide to writing what must be written: proposals, literature reviews, theses, journal articles, slide presentations, posters, and grant applications. The author also discusses conventions in writing, proofreading and copywriting, as well as methods for searching and citing scientific literature, composing reviews, preparing data presentations, communicating visually and public speaking.

Fink, A. (1995) *How to Report on Surveys* (The Survey Kit, Vol. 9). Thousand Oaks, CA: Sage Publications.

*How to Report on Surveys* demonstrates how to prepare more effective verbal and written survey reports. The book includes guidelines for preparing overhead transparencies and a step-by-step account of how to talk about a survey and the data analysis. It also shows how to prepare written reports for both technical and general audiences.

Gall, J. P., Gall, M. D. and Borg, W. R. (Eds.) (1999) *Applying Educational Research: A Practical Guide* (4th ed.). New York: Longman.

Designed for courses focused solely on consuming rather than conducting research, this textbook includes 13 actual research articles, reprinted in their entirety. The primary author of each article then offers original commentary on his/her piece. Through this format, the textbook presents a comprehensive explanation of the methodologies used by present-day researchers, data-collection challenges, and the meaning of the results. Part 3, 'Quantitative Research in Education', includes chapters on reading reports of quantitative research studies, statistical analysis of research data, descriptive and causal-comparative research, and correlational research and experimental research.

Henry, G. T. (Ed.) (1997) *Creating Effective Graphs: Solutions for a Variety of Evaluation Data, New Directions for Evaluation*. San Francisco, CA: Jossey-Bass.

The chapters in this sourcebook describe researchers' responses to problems of data presentation. Each author describes the evaluation problem, the data and the software employed for the final graph. Besides giving many examples of well-composed and well-executed graphs, the authors explain the role of graphs in the communication of findings.



Huck, S. W. (2000) *Reading Statistics and Research* (3rd ed.). New York: Longman.

This textbook explains to consumers of research how to read, understand and critically evaluate the statistical information contained in technical research reports.

Kumar, R. (1999) *Research Methodology: A Step-by-Step Guide for Beginners*. Thousand Oaks, CA: Sage Publications.

This practical book is designed for people new to research and planning who intend to undertake a research project. The textbook is organized around eight operational steps that constitute the research process: formulating a research problem; conceptualizing a research design; constructing an instrument for data collection; selecting a sample; writing a research proposal; collecting data; processing data; and writing a research report. Chapter 16, 'Displaying Data', and Chapter 17, 'Writing a Research Report', provide an easy-to-read coverage of the major points in report writing.

Lester, J. D. (1999) *Writing Research Papers: A Complete Guide* (9th ed.). New York: Longman.

This book provides step-by-step guidance through the research writing process from selecting and narrowing a topic to formatting the finished document.

Lester, J. D. and Lester, J. D. Jr. (1999) *An Essential Guide to Writing Research Papers*. New York: Longman.

This is a more concise version of J. D. Lester's *Writing Research Papers: A Complete Guide*. The book covers all aspects of the research-writing process from selecting a topic and gathering data to formatting the final draft. However, in the interests of brevity, the authors make more frequent use of checklists and summaries than is the case in *Writing Research Papers*.

Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A. and O'Connor, K. M. (2000) *Third International Mathematics and Science Study: International Science Report*. Chestnut Hill, MA: Boston College.

This international report of the results of the TIMSS 1999 eighth-grade science assessment provides an example of the current TIMSS approach to presenting findings. Written in non-technical language, the book is intended for use by policy-makers, educators and other interested readers.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J. and Smith, T. A. (2000) *Third International Mathematics and Science Study: International Mathematics Report*. Chestnut Hill, MA: Boston College.

This international report of the results of the TIMSS 1999 eighth-grade mathematics assessment provides an example of the current TIMSS approach to presenting findings. It is written in non-technical language, and is intended for use by policy-makers, educators and other interested readers.



Paradis, J. G. and Zimmerman, M. (1997) *The MIT Guide to Science and Engineering Communication*. Cambridge, MA: MIT Press.

*The MIT Guide* emphasizes processes and forms that will help in creating documents, and includes numerous examples. There is also a strong emphasis on the central role of the computer in creating and disseminating technical materials.

Torney-Purta, J., Lehmann, R., Oswald, H. and Schulz, W. (2001) *Citizenship and Education in Twenty-eight Countries*. Amsterdam: IEA.

This international report from the IEA Civic Education Study is a good example of a report of a recent IEA project.



## Standard for Reviewing the Primary Reports of Study Findings

### Purpose

To ensure that the primary reports of study findings present the most important findings accurately and in a way that is consistent with the aims of the study. The publication plan for the study as well as primary reports of study findings should be thoroughly reviewed by the appropriate constituencies and advisory committees before publication.

### Standard

The publication plan for the study as well as primary reports of study findings must be thoroughly reviewed by the appropriate constituencies and advisory committees before publication. These reviews should occur at critical points during report development and should comprise, for example, the report outline, shells for the proposed tables and graphs, drafts of the text and the data for tables and graphs, and the final draft of the complete report.

Depending on the resources available for the study, the reviews should be conducted in meetings of the relevant parties so that differing points of view can be resolved in the best way possible. However, supplementary teleconferencing and mail reviews also will be necessary. All steps of the review process should be conducted in accordance with a schedule prepared well in advance of the publication deadline, and all parties involved in the review should be given adequate notification about the review schedule.

### Summary of the Guidelines

The publication plan should be reviewed by all interested parties, including subject-matter advisory committees, the technical advisory committee, and the IEA publications and editorial committee, as well as by national research coordinators (NRCs) and their advisors. Report outlines and draft data tables should be reviewed by NRCs early in the report-writing process, so that changes may be incorporated as necessary. NRCs should review drafts of the reports, including both data tables and text, before these are finalized. The publications and editorial committee should review the final drafts.

### Source for this Standard

McMillan, J. H. (2000) *Educational Research: Fundamentals for the Consumer* (3rd ed.). New York: Longman.

This book is aimed at university students, and seeks to make them intelligent consumers of education research. The section 'Anatomy of a Research Article' explains excerpts from recent research studies so as to show students the function of various features of research. Chapters 6 and 7, on educational measurement, include the latest research on reliability and validity.



## Standard for Releasing Data

### Purpose

To ensure that an international database is available under pre-defined conditions.

### Standard

A data release policy should be decided at the beginning of the study. Agreement with this policy should be secured in writing from all participants at an early stage. Participants should also agree on a timetable that defines when and for whom the data are available and whether a distinction should be made between users who are from institutions involved with IEA and users from other institutions. At the end of the study, the released data should be made available to all interested users.

### Summary of the Guidelines

A data release policy should be decided at the beginning of the study. Agreement with this policy should be secured in writing from all participants at an early stage. Participants should be given the opportunity to discuss and agree on timelines for the data release and the format for the data release.

### Sources for this Standard

Gonzalez, E. J. and Smith, T. A. (Eds.) (1997) *User Guide for the TIMSS International Database: Primary and Middle School Years*. Chestnut Hill, MA: Boston College.

This volume and four accompanying supplementary volumes describe the TIMSS 1995 database for primary and middle school grades (Grades 3 and 4 and Grades 7 and 8 in most countries) and provide instructions for accessing and analyzing the data.

Gonzalez, E. J., Smith, T. A. and Sibberns, H. (Eds.) (1998) *User Guide for the TIMSS International Database: Final Year of Secondary School*. Chestnut Hill, MA: Boston College.

This volume and three accompanying supplementary volumes describe the TIMSS 1995 database for the final year of secondary school (Grade 12 in most countries) and provide instructions for accessing and analyzing the data.

Gonzalez, E. J. and Miles, J. A. (Eds.) (2001) *TIMSS 1999 User Guide for the International Database: IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.

This volume and three accompanying supplementary volumes describe the TIMSS 1999 database, which contains data from both the 1995 and 1999 TIMSS eighth-grade assessments. The volume describes the structure and contents of the database, and provides instructions and program segments to assist in accessing and analyzing the data.





Pascal, F. (1993) *Understanding Rational Databases with Examples in SQL-92*.  
New York: John Wiley and Sons.

This book examines the fundamentals of relational concepts and then demonstrates the applications of these concepts. The author provides numerous practical examples using structured query language (SQL).



## Standard for Preparing Technical Reports and Documentation

### Purpose

To ensure that the study design, instrument development, data collection, and analysis and reporting procedures are described in sufficient detail to permit them to be evaluated and replicated.

### Standard

All aspects of the study design, development, operations and analysis should be clearly documented in a timely manner. Reports intended for a general audience should include a brief technical appendix with an explanation of the major technical features of the study. Detailed technical documentation aimed at a technically sophisticated audience should also be produced, ideally in a separate technical report. This document should include specifications of all design, procedural and analytic aspects of the study.

### Summary of the Guidelines

All aspects of the study design, development, operations and analysis should be clearly documented in a timely manner.

### Sources for this Standard

Martin M. O. and Kelly, D. L (Eds.) (1996) *Third International Mathematics and Science Study Technical Report. Volume 1: Design and Development*. Chestnut Hill, MA: Boston College.

This volume documents the design and development activities for TIMSS 1995. More specifically, it includes design and development of the achievement tests and questionnaires; performance assessment; translation and cultural adaptation; field operations training for scoring free response achievement questions; and procedures for quality control. Individual chapters of this volume are referenced throughout this bibliography, as appropriate.

Martin M. O. and Kelly, D. L (Eds.) (1997) *Third International Mathematics and Science Study Technical Report. Volume 2: Implementation and Analysis: Primary and Middle School Years*. Chestnut Hill, MA: Boston College.

This volume documents the implementation and analysis of TIMSS 1995 at the primary and middle school grades (Grades 3 and 4 and Grades 7 and 8 in most countries). Chapters cover the computing sampling weights and standard errors, scaling the achievement data and constructing the databases, and reporting the mathematics and science achievement and questionnaire data.



Martin M. O. and Kelly, D. L (Eds.) (1998) *Third International Mathematics and Science Study Technical Report. Volume 3: Implementation and Analysis: Final Year of Secondary School*. Chestnut Hill, MA: Boston College.

This volume documents the implementation and analysis of TIMSS 1995 at the final year of secondary school (Grade 12 in most countries). Chapters cover the computing sampling weights and standard errors, scaling the achievement data and constructing the databases, and reporting the mathematics and science achievement and questionnaire data.

Martin, M. O., Gregory, K. D. and Stemler, S. E. (Eds.) (2000) *TIMSS 1999 Technical Report: IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.

This report describes all of the technical activities underpinning TIMSS 1999, including test and questionnaire development; sampling design; field operations; quality control; scaling, weighting and database construction; and the analysis and reporting of the mathematics and science achievement data.

Martin, M. O. and Mullis, I. V. S. (Eds.) (1996) *Third International Mathematics and Science Study Quality Assurance in Data Collection*. Chestnut Hill, MA: Boston College.

This volume describes the program of international quality assurance activities that was implemented as part of the TIMSS 1995 data collection. It includes chapters on translation verification, sampling, test administration and data collection, reliability of free-response scoring, and data consistency checking across countries.

Whitaker, K. (1995) *A Guide to Publishing User Manuals*. New York: John Wiley and Sons.

Writing for people with little or no desktop publishing experience, the author guides the reader through the entire process, from organizing the material and writing the documentation to designing and printing the manual. Whitaker discusses planning and writing the manual, as well as a range of presentation considerations and techniques.

